



***UNIVERSIDAD NACIONAL DE LA PLATA***

***FACULTAD DE CIENCIAS EXACTAS***

***DEPARTAMENTO DE CIENCIAS BIOLÓGICAS***

---

Trabajo de Tesis Doctoral

**Desarrollo de herramientas moleculares para  
la evaluación de la calidad genética y  
productividad en la cría artificial de  
*Diachasmimorpha longicaudata*, agente de  
control biológico de moscas plaga de los frutos**

---

Lic. Mannino Maria Constanza

Directora: Dra. Lanzavecchia Silvia

Co-Director: Dr. Romanowski Víctor

2016

El presente trabajo de tesis para optar por el grado de Doctor de la Facultad de Ciencias Exactas de la Universidad Nacional de La Plata, ha sido realizado bajo la dirección de la Dra. Lanzavecchia Silvia en el Instituto de Genética “Ewald A. Favret” (INTA) y la co-dirección del Dr. Romanowski Víctor en el Instituto de Biotecnología y Biología Molecular (IBBM, CONICET – UNLP).

La Plata, Marzo de 2016.

*“...quizás tendría que negociar con el gusano de seda por el elástico de mis pantalones...”*

*(Abogado defensor de las corporaciones que comercializan miel, Bee Movie 2007)*

*Muchos animales trabajan para los humanos, los insectos no son la excepción...*

*A ellos...*

## Agradecimientos

Este trabajo de Tesis es el producto del esfuerzo conjunto a través del tiempo de un conjunto de personas a las cuales les estoy profundamente agradecida.

A mi directora la Dra. Lanzavecchia y mi co-director el Dr. Romanowski por su tiempo y dedicación. Al Dr. Cladera por permitirme trabajar con él y aportar siempre su valiosa experiencia.

Agradezco a la UNLP, en especial a todos los docentes dedicados de la Facultad de Ciencias Exactas por la formación adquirida. A la ANPCyT y al CONICET por otorgarme las becas que me permitieron formarme como profesional y desarrollar mi doctorado.

A todos los integrantes del IGEAF, quienes han colaborado con mi formación en diferentes aspectos. A la Dra. Farber, Dr. Rivarola e Ing. Gonzalez del Instituto de Biotecnología perteneciente al CICVyA, por su colaboración en el área de Bioinformática, predisposición a consultas y aportes sin los cuales este trabajo no hubiera sido posible. Mi agradecimiento a los miembros del laboratorio 2 del IBBM, en especial a la Dra. Ferrer por su excelente predisposición.

Un agradecimiento especial a la Dra. Scannapieco, Ale, por la invaluable ayuda, compañía y paciencia durante todo este camino. A Clau, Mari, Leo, Ceci, Faby, Fran, Guille, Diego, Clari, Ana, Iri y Germán compañeros del Laboratorio de Insectos de Importancia Económica por permitirme aprender con ellos a lo largo de estos años.

A mis amigos que a lo largo de estos años han sido oídos y compañía en todos los momentos, a Lety y Cos por el aguante continuado. A Pao, Mari y Ale.

A mi familia que me acompañó y contuvo en todo este proceso. A mis viejos y a mi hermano por el cariño, por haberme dado siempre todas las posibilidades y el apoyo incondicional. A Maxi, mi chancho, en el último tramo de este camino y en el inicio de muchos más. Sin ustedes no sería la persona que soy.

A todos... simplemente Muchas Gracias!



## **Abreviaturas**

**A:** Adenina

**ADN:** ácido desoxirribonucleico

**ADNc:** ácido desoxirribonucleico complementario

**ADNsas:** Enzimas de degradación de ADN

**ARN:** ácido ribonucleico

**ARNsas:** Enzimas de degradación de ARN

**ARNm:** ácido ribonucleico mensajero

**ARNr:** ácido ribonucleico ribosomal

**BLAST:** Herramienta de búsqueda de alineamientos locales básicos (“Basic Local Alignment Search Tool”).

**C:** Citosina

**CB:** Control Biológico

**G:** Guanina

**GO:** Ontología Génica

**IBBM:** Instituto de Biotecnología y Biología Molecular

**IGEAF:** Instituto de Genética “Ewald A. Favret”

**INDEAR:** Instituto de Agrobiotecnología Rosario

**INTA:** Instituto Nacional de Tecnología Agropecuaria

**MK:** Marcador de peso molecular

**NGS:** secuenciación de próxima generación

**PCR:** reacción en cadena de la polimerasa

**qPCR:** reacción en cadena de la polimerasa en tiempo real (cuantitativa)

**RT:** transcripción reversa

**SNP:** Polimorfismos de Nucleótido Simple (“Single Nucleotide Polymorphism”)

**SSR:** Repeticiones de secuencia simple (“Simple Sequence Repeat”)

**T:** Timina

## RESUMEN

*Diachasmimorpha longicaudata* Ashmead (Hymenoptera: Braconidae) es un endoparásitoide solitario de estadios larvales de moscas de la fruta perteneciente a la familia Tephritidae. Es criado a nivel masivo en bioplasmas y utilizado en diversas partes del mundo para las estrategias de control biológico (CB) principalmente de dípteros de importancia económica de los géneros *Ceratitis*, *Anastrepha* y *Bactrocera*. Actualmente, se estudia su implementación en nuestro país para el control de *Ceratitis capitata* Wiedemann (Diptera: Tephritidae). Se cuenta con una vasta cantidad de información acerca de su biología general, fisiología, comportamiento y condiciones de cría artificial. Recientemente se han descrito aspectos genéticos y citogenéticos que permiten profundizar los conocimientos respecto a sus procesos fisiológicos. Este parasitoide posee un ciclo de vida haplodiploide (machos haploides, hembras diploides) y la determinación sexual está gobernada por el mecanismo de alelos complementarios CSD (de las siglas en inglés “Complementary sex determination”) con al menos dos *loci* involucrados en las vías tempranas de este complejo mecanismo, aunque aún son escasos los conocimientos que se poseen acerca de sus bases genéticas. Estos conocimientos tienen un impacto potencial relevante tanto sobre el diseño de estrategias de mejoramiento de la productividad y calidad de las crías artificiales como sobre la implementación del CB basado en *D. longicaudata* en el campo.

A fin de aportar información genética para mejorar las condiciones artificiales de cría de *D. longicaudata* y profundizar el conocimiento sobre aspectos moleculares de mecanismos de relevancia para la aplicación del CB en nuestro país, este trabajo de tesis se centró en el estudio de las bases moleculares del sistema de determinación del sexo en esta especie. Se tomaron dos aproximaciones, una aproximación indirecta, caracterizando regiones genómicas a partir del conocimiento de genes involucrados en la determinación del sexo en especies relacionadas y una aproximación directa mediante el análisis de regiones codificantes específicas a través de la secuenciación del transcriptoma de la especie y el análisis comparativo descriptivo de expresión entre machos y hembras del parasitoide. A partir de esta última aproximación, se generó una cantidad masiva de datos de alta calidad, que permitieron conocer una alta proporción de las secuencias expresadas en *D. longicaudata*. Mediante una búsqueda por homología de secuencia a partir de nuestra base de datos, se ha logrado identificar transcritos homólogos a dos de las llaves involucradas en la determinación del sexo: el regulador transcripcional *feminizer/transformer* y el efector que se encarga de regular genes que determinarán caracteres sexuales secundarios, *doublesex*. Se validaron *a posteriori* los resultados por medio de la técnica de RT-qPCR y se analizó la expresión diferencial de las variantes de *fem* y *dsx* entre machos y hembras; encontrando que existen dos variantes del transcripto homólogo a *fem* que se

expresan diferencialmente entre sexos. Paralelamente, la tecnología de secuenciación masiva permitió la identificación y caracterización *in silico* de 7000 marcadores moleculares potenciales (SSR y SNP) estableciendo la utilidad de 1106 marcadores microsatélites disponibles para estudios genético-poblacionales para caracterizar la variabilidad genética de poblaciones silvestres y cepas establecidas en condiciones de laboratorio o cría masiva de *D. longicaudata*.

**Palabras clave:**

*Diachasmimorpha longicaudata*; determinación del sexo; genes candidatos; transcriptómica; expresión diferencial, marcadores moleculares

# Índice

<b>OBJETIVOS .....</b>	<b>11</b>
Objetivo General .....	11
Objetivos Particulares.....	11
<b>INTRODUCCIÓN.....</b>	<b>12</b>
El orden Hymenoptera .....	13
1. El orden Hymenoptera .....	14
1.1. Taxonomía del orden Hymenoptera .....	14
1.2. Importancia de los himenópteros parasíticos .....	16
<i>Diachasmimorpha longicaudata</i> (Hymenoptera: Braconidae) .....	17
2. <i>Diachasmimorpha longicaudata</i> (Ashmead) (Hymenoptera: Braconidae) .....	18
2.1. Biología general de la especie .....	18
2.2. <i>D. longicaudata</i> como agente de control biológico .....	19
2.3. <i>D. longicaudata</i> en Argentina .....	20
2.4. Genética de <i>D. longicaudata</i> .....	21
Mecanismos de determinación del sexo .....	23
3. Determinación del sexo.....	24
3.1. Determinación del sexo en insectos .....	25
3.2. Determinación del sexo en Hymenoptera (antecedentes principales en <i>A. mellifera</i> y <i>N. vitripennis</i> ) .....	25
3.3. Determinación del sexo en <i>D. longicaudata</i> .....	26
Tecnologías de secuenciación masiva: transcriptoma asociado al sexo.....	27
4. Pirosecuenciación 454: aplicaciones.....	28
4.1. Aplicaciones de la pirosecuenciación: descubrimiento de genes y marcadores .....	30
<b>RESULTADOS .....</b>	<b>32</b>
Extracción de ARN y búsqueda de Genes Candidatos.....	33
1. Identificación de genes involucrados en la determinación del sexo por homología .....	34
1.1. Puesta a punto de la extracción de ARN .....	34
1.1.1. Volumen de Trizol en relación a la masa de tejido de partida.....	35
1.1.2. Con y sin uso de ADNasa I.....	37

1.1.3.	Puesta a punto de la Transcripción Reversa .....	39
1.2.	PCRs inespecíficas de los Genes de Interés: <i>csd</i> y <i>dsx</i> .....	41
1.2.1.	Gen <i>csd</i> .....	41
1.2.2.	Gen <i>dsx</i> .....	42
	Análisis del Transcriptoma de <i>D. longicaudata</i> .....	47
2.	Transcriptómica de <i>D. longicaudata</i> : Identificación de genes por Pirosecuenciación 454 .....	48
2.1.	Caracterización del transcriptoma de <i>D. longicaudata</i> ....	49
2.1.1.	Obtención de las muestras de ARN .....	49
2.1.2.	Pirosecuenciación 454 y ensamblado del transcriptoma de <i>D. longicaudata</i> .....	50
2.2.	Obtención de marcadores moleculares a partir de ARN .	55
2.2.1.	Diversidad Genética: predicción de marcadores.....	55
2.2.1.1.	Obtención <i>in silico</i> de Repeticiones de Secuencia Única (SSR) .....	56
2.2.1.2.	Caracterización de motivos y distribución de microsatélites .....	56
2.3.	Anotación funcional: identificación y análisis de secuencias .....	58
	Determinación del sexo en <i>D. longicaudata</i> .....	65
3.	Análisis de Genes Candidatos .....	66
3.1.	Validación de resultados de secuenciación .....	66
3.1.1.	Puesta a punto y determinación del rango de linealidad de la RT-qPCR: Genes de Referencia y Genes Candidato .....	66
3.1.1.1.	Puesta a punto de la RT-qPCR y construcción de las curvas de calibración .....	66
3.1.1.1.1.	Elección de GR .....	66
3.1.1.1.2.	Elección de GI .....	67
3.1.1.2.	Curvas de calibración .....	67
3.1.2.	Validación de los patrones de expresión por medio de análisis comparativo .....	69
3.2.	Análisis comparativo del transcriptoma .....	71
3.3.	Selección de genes de determinación del sexo .....	72
3.3.1.	Caracterización de transcritos previamente obtenidos: <i>csd</i> .....	73
3.3.2.	Caracterización de secuencias de genes candidato identificados por transcriptómica .....	73
3.3.2.1.	<i>Feminizer</i> .....	74

3.3.2.1.1.	Alineamiento general para las variantes de feminizer.....	75
3.3.2.2.	<i>Doublesex</i> .....	78
3.4.	Análisis de expresión diferencial entre sexos: genes de determinación del sexo .....	81
3.4.1.	Expresión diferencial en adultos .....	81
3.4.2.	Expresión diferencial a lo largo del desarrollo .....	84
3.4.2.1.	Obtención de muestras y establecimiento de modificaciones a los protocolos .....	84
3.4.2.1.1.	Relación de masa y cantidad de individuos vs estadio del desarrollo .....	84
3.4.2.2.	Cuantificación a lo largo del desarrollo: validación de los Genes de Referencia .....	85
3.4.2.3.	Patrones de expresión para los GI a lo largo del desarrollo .....	85
<b>DISCUSIÓN GENERAL .....</b>		<b>89</b>
<b>CONCLUSIONES GENERALES .....</b>		<b>100</b>
<b>MATERIALES Y MÉTODOS .....</b>		<b>102</b>
1.	Insectos.....	103
2.	Extracción de ADN.....	103
3.	Extracción de ARN.....	103
4.	Tratamientos con ADNasa I .....	104
5.	Transcripción Reversa .....	104
6.	Selección de Genes de Interés (GI) .....	105
7.	PCR de punto final.....	105
8.	Purificación .....	106
9.	Secuenciación Sanger .....	106
10.	Análisis de Secuencias.....	106
11.	Pirosecuenciación 454.....	106
12.	Ensamblado y anotación del transcriptoma.....	107
13.	Obtención de potenciales marcadores moleculares: SSR y SNP .....	108
14.	Identificación de transcritos diferencialmente expresados .....	108
15.	Selección de GI.....	109
16.	qPCR .....	111
<b>BIBLIOGRAFÍA .....</b>		<b>112</b>
<b>ANEXOS .....</b>		<b>124</b>

# OBJETIVOS

## OBJETIVO GENERAL

Este proyecto enmarca el estudio a nivel genético y molecular de *Diachasmimorpha longicaudata* (Hymenoptera: Braconidae) el controlador biológico más utilizado en estrategias de control de bajo impacto ambiental de moscas plaga de la fruta de la familia Tephritidae. Este trabajo busca profundizar el conocimiento sobre las bases moleculares de la determinación del sexo, así como también generar herramientas para el análisis de la variabilidad genética en la cría artificial del parasitoide con la finalidad de aportar información que contribuya a mejorar el rendimiento y asegurar la calidad genética de los insectos producidos, aportando así herramientas de utilidad para el desarrollo de la estrategia de control biológico de moscas de los frutos en nuestro país mediante el uso de insectos parasitoides.

## OBJETIVOS PARTICULARES

1. Caracterización de genes involucrados en el mecanismo de determinación del sexo en *Diachasmimorpha longicaudata*.
  1. a. Caracterización de genes involucrados en la determinación del sexo por medio de la búsqueda de homología con especies relacionadas.
  1. b. Búsqueda de genes involucrados en la determinación del sexo a partir de la caracterización del transcriptoma de *D. longicaudata*.
  1. c. Caracterización de los perfiles de expresión entre sexos. Expresión de *feminizer* y *doublesex*.
2. Desarrollo de potenciales marcadores moleculares específicos para *D. longicaudata* para el establecimiento de parámetros de calidad genética en la cría experimental de *D. longicaudata* cepa IGEAF-INTA.

# INTRODUCCIÓN



# El orden Hymenoptera

## 1. El orden Hymenoptera

### 1.1. Taxonomía del orden Hymenoptera

El orden Hymenoptera comprende aproximadamente el 10% de la diversidad de especies del planeta y es considerado uno de los cuatro órdenes megadiversos de insectos junto a Coleoptera, Diptera y Lepidoptera (Whitfield, 1998). Presenta entre 115.000 y 199.000 especies distribuidas entre 79 y 99 familias (Goulet and Huber, 1993; Hanson and Gauld, 1995; Fernández and Sharkey, 2006). La familia más numerosa dentro del orden es la Icheumonidae que contiene una cantidad de miembros que superan en número a todas las especies conocidas de aves y mamíferos.

Los himenópteros poseen un rol fundamental en los ecosistemas debido a sus diversas funciones e interacciones con otros seres vivos y a sus implicancias ecológicas y económicas, por este motivo es vital su conservación y estudio. La composición del orden es muy diversa, contando entre sus miembros especies perjudiciales o plaga como las hormigas; polinizadores de gran importancia como abejas y abejorros; y especies con potencial para aplicaciones biotecnológicas como son los controladores biológicos (parasitoides) (Wilson, 1987; LaSalle and Gauld, 1993).

Las principales características de este orden comprenden un ciclo de vida holometábolo (huevo-larva-pupa-adulto), dos pares de alas membranosas transparentes, un aparato bucal de tipo masticador, la presencia de un ovipositor típico y homólogo al de los insectos más primitivos y una diferenciación de regiones de cuerpo o tagmas particular, lo que dificulta relacionarlos claramente con el resto de los órdenes de holometábolos. En cuanto a su forma de vida, existen muchas diferencias dentro del grupo, encontrándose tanto especies parásitas como de vida libre (Huber, 2009).

Las especies de Hymenoptera se dividen en dos subórdenes: Symphyta y Apocrita. Los apócritos se subdividen a su vez en dos grupos, Acuelata y Parasítica. Mientras que dentro del grupo Acuelata pueden incluirse las hormigas, abejas y avispas (el ovipositor de la hembra se transforma en un aguijón), los himenópteros que pertenecen al grupo Parasítica tienen una forma de vida parásita durante sus estadios inmaduros y son adultos de vida libre. Este tipo de himenópteros se denominan parasitoides y el término responde a que en su mayoría el ciclo de vida implica la muerte del hospedador. Usualmente tienen un tamaño similar a su hospedador y pertenecen al mismo taxón (Martínez, 2009).

Los parasitoides pertenecientes al orden Hymenoptera pueden clasificarse en: ecto o endoparasitoides; idiobiontes o koinobiontes; solitarios o gregarios y de huevo, larva, pupa o adulto según sus interacciones con el hospedador y con otros parasitoides. Las formas ectoparasíticas se desarrollan alimentándose de los tejidos del hospedador desde el exterior; los endoparasitoides, en cambio, se alimentan del tejido del hospedador desde el interior del mismo. Estas categorías, sin embargo, no son absolutas, pudiendo un mismo parasitoide iniciar su ciclo de vida como endoparasitoide y continuar alimentándose el hospedador una vez emergido (Debach and Rosen, 1991). La clasificación en idiobiontes y koinobiontes se refiere al efecto del ciclo de vida del parasitoide sobre el hospedador. Los idiobiontes suelen ser ectoparasitoides e inyectan un veneno en el hospedador en el momento de la oviposición que afecta significativamente su desarrollo. Los parasitoides koinobiontes no inyectan veneno y se alimentan y desarrollan en paralelo con su hospedador, el que morirá cuando éste haya consumido masivamente sus órganos internos (DeBach and Rosen, 1991). Los parasitoides solitarios se desarrollan individualmente dentro del hospedador y los parasitoides gregarios pueden desarrollarse en grupo dentro del mismo hospedador (Martínez, 2009). Algunos grupos son muy uniformes con respecto al estadio del hospedador atacado. En general, hay pocos parasitoides del estado adulto de hospederos holometábolos, en cambio, son muy comunes los parasitoides de larvas (Hanson and Gauld, 1995).

Dentro del grupo Parasítica se incluyen las superfamilias Ceraphronoidea, Chalcidoidea, Cynipoidea, Evanioidea, Ichneumonoidea, Megalyroidea, Mymarommatoidea (a veces incluida en Serphitoidea), Platygastroidea, Proctotrupeoidea, Serphitoidea (fósil), Stephanoidea y Trigonoidea. De particular interés para la presente Tesis Doctoral, es la superfamilia Ichneumonoidea, que incluye sólo dos familias: Ichneumonidae y Braconidae. En esta última superfamilia se han descrito más de 80.000 especies, siendo la gran mayoría controladores biológicos de especies de insectos plaga. En muchas de estas especies de parasitoides se ha descrito la presencia de un polyADNvirus posiblemente involucrado en la represión del sistema inmunitario del hospedador facilitando la oviposición exitosa (Lawrence, 2005a, 2005b). La familia Braconidae es la segunda en cuanto a diversidad específica dentro del orden Hymenoptera (Shaw, 1995) y puede distinguirse por la siguiente combinación de caracteres morfológicos: ala anterior con la vena 2m-cu (segunda vena recurrente) ausente, por lo que la segunda celda discal se encuentra abierta; tergos II y III del metasoma fusionados; el ala anterior con la vena Rs+M generalmente presente, separando la primera celda submarginal de la primera celda discal; el ala posterior con la vena 1r-m basal respecto a la separación de las venas R1 y Rs (Martínez, 2009).

## 1.2. Importancia de los himenópteros parasíticos

Las especies comprendidas en el orden Hymenoptera, grupo Parasítica modifican indirectamente la estructura del ecosistema al afectar el tamaño poblacional de ciertas especies de insectos herbívoros, es decir, poseen el rol de controladores biológicos en los ecosistemas (Lasalle and Gauld, 1991). Las implicancias económicas que se desprenden directamente son muy amplias, ya que existen muchas especies de insectos fitófagos que afectan cultivos de interés económico y generan anualmente pérdidas millonarias (Suarez et al., 2014). En este sentido, y utilizando la función biológica de estas especies, este grupo de insectos himenópteros se cría en forma masiva y se libera al campo como parte de programas de control integrado de plagas de bajo impacto ambiental, en conjunto con otras medidas biológicas, biotecnológicas, químicas y comportamentales a fin de minimizar el empleo de insecticidas para el control de poblaciones plaga. El método que incluye la liberación de enemigos naturales como parasitoides, predadores, patógenos, antagonistas o competidores de una plaga, se denomina Control Biológico (CB) y tiene como finalidad suprimir y/o reducir su tamaño poblacional por debajo del umbral de daño económico (Ovruski and Schliserman, 2012; Suarez et al., 2014).

Otro factor de importancia a tener en cuenta cuando se habla de los himenópteros braconidos particularmente, es que tienen una potencial aplicación como bioindicadores por la relación que establecen con su hospedador. Esta particularidad permitiría estimar el tamaño de las poblaciones de insectos fitófagos siempre teniendo en cuenta el conocimiento y estabilidad de la taxonomía del grupo y la comprensión de la biología e historia de vida de los organismos involucrados (Favila and Halffter, 1997; Pearson and Cassola, 1992).

*Diachasmimorpha longicaudata*  
(Hymenoptera: Braconidae)

## 2. *Diachasmimorpha longicaudata* (Ashmead) (Hymenoptera: Braconidae)

### 2.1. Biología general de la especie

*Diachasmimorpha longicaudata* es un endoparasitoide koinobionte solitario de moscas plaga de la fruta. Esta especie es nativa de la región indomalaya (Bess, 1961; Clausen et al., 1965). Las hembras adultas depositan sus huevos en estadios larvales tardíos de moscas de los frutos de los géneros *Bactrocera* (Macquart), *Anastrepha* (Schiner) y *Ceratitis* (Macleay) (Diptera, Tephritidae), aunque algunos estudios indican que pueden desarrollarse ocasionalmente en larvas de otros dípteros (Terán López, 1983).



**Figura 1.** Hembra adulta de *D. longicaudata* sobre sustrato artificial de oviposición.

El ciclo de vida de *D. longicaudata* se inicia en el estado de huevo y consta de tres estadios larvarios, pupa y adulto (Carabajal Paladino et al., 2010). Desde el huevo hasta la pupa se desarrolla en el interior del hospedador y el adulto es de vida libre (Greany et al., 1976; Pemberton, CE; Willard, 1918). La duración de cada uno de los estadios es altamente dependiente de las condiciones ambientales durante el desarrollo, así como la edad de las larvas del hospedador (Lawrence et al., 1976). En condiciones de laboratorio con temperaturas de  $25 \pm 1$  °C, humedad de 50 - 70 % y fotoperíodos de 12:12 horas luz/oscuridad, la emergencia de los individuos adultos macho se produce alrededor de 15 días después de la oviposición, mientras que las hembras emergen 24 a 48 horas después (Carabajal Paladino et al., 2010).

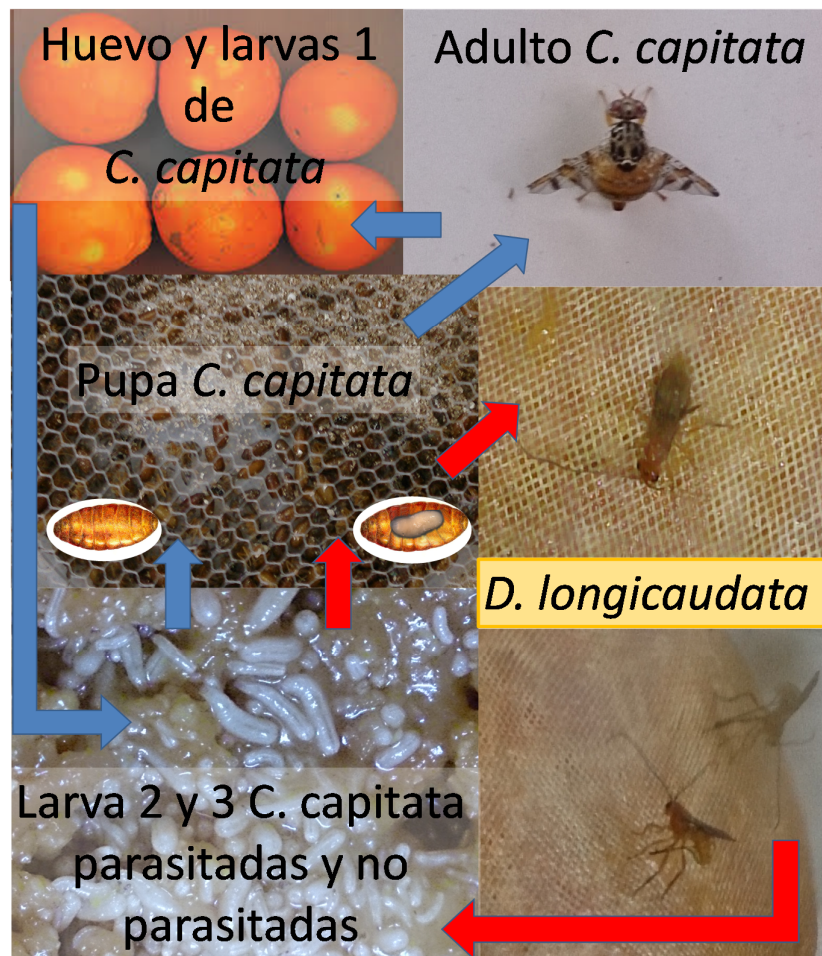
Al ser un parasitoide solitario, de cada larva parasitada emerge un solo parasitoide adulto originado de un único huevo depositado. Existen evidencias de superparasitismo (presencia de más de un estadio inmaduro del parasitoide por hospedador) en condiciones de laboratorio con emergencia de un único parasitoide adulto (Devescovi, 2015; Greany et al., 1976). Luego de la emergencia, de manera inmediata se produce la cópula breve, con una duración de escasos segundos, la hembra puede copular con más de un macho a lo largo de su vida (Hagen, 1953; Martinez, 1992). Si bien la oviposición puede ocurrir inmediatamente después de la emergencia de los parasitoides adultos, es más frecuente que ocurra a partir de los nueve días posterior a la misma (Vargas et al., 2002; Viscarret et al., 2006b).

La atracción del parasitoide hacia los hospedadores se produce por varios mecanismos: detección de compuestos volátiles (como acetaldehídos) emitidos por las frutas en descomposición y generados por acción de las larvas del hospedador; daño externo de los frutos; vibraciones producidas por las larvas del hospedador en el interior de las frutas y sonidos que esta realiza durante su alimentación (Greany et al., 1977; Lawrence, 1981; Segura et al., 2012). El parasitoide logra localizar la larva en el interior de la fruta a través de sus antenas y los mecanorreceptores ubicados en los tarsos de las patas anteriores (Leyva et al., 1988) y en el ovipositor (Devescovi, 2015). El parasitoide realiza pruebas con el ovipositor para evaluar el estado fisiológico (censando la composición de hormonas en la hemolinfa) y tamaño de la larva del hospedador (Martínez, 1989). Estas pruebas también le indican a la hembra si la larva de hospedador ha sido parasitada o no (Lawrence, 1982).

## 2.2. *D. longicaudata* como agente de control biológico

*D. longicaudata* tiene relevancia económica por ser ampliamente utilizado como parte integral de programas de Control Biológico de moscas plaga de la fruta de la familia Tephritidae que causan importantes pérdidas económicas anualmente (Guillén and Sánchez, 2007). La cría artificial de este parasitoide es relativamente sencilla. Respecto a las condiciones de campo, posee gran habilidad para adaptarse a los ecosistemas en donde es liberado y posee un amplio rango de hospedadores, estas características hacen que sea un excelente candidato para llevar a cabo el CB (Marucci and Clancy, 1950). Esta especie de parasitoide ha demostrado hospedarse eficientemente en *Anastrepha fraterculus* (Wiedemann) (Diptera: Tephritidae) así como en *Ceratitis capitata* (Wiedemann) (Diptera: Tephritidae) tanto en regiones tropicales (Cancino et al., 1995; Sivinski, 1996) como en ambientes áridos (Greany et al., 1976). Adicionalmente, en Hawaii (Purcell, 1998; Vargas et al., 2012) y México (González et al., 2007; Montoya et al., 2007, 2012) se han desarrollado protocolos para la cría masiva con la finalidad de realizar liberaciones

aumentativas del parasitoide que han sido implementados con éxito (Schliserman et al., 2003a; Suarez et al., 2014).



**Figura 2.** Ciclo de vida de *D. longicaudata* en condiciones de laboratorio. Se indica con flechas azules el ciclo de vida de *C. capitata* y con flechas rojas el ciclo de vida de *D. longicaudata*.

### 2.3. *D. longicaudata* en Argentina

En nuestro país, la utilización de este parasitoide podría ser útil en áreas donde *C. capitata* y *A. fraterculus* están presentes, situación que ocurre en las regiones citricolas del Norte y Centro de la Argentina, particularmente Jujuy, Salta, Tucumán y Catamarca (Suarez et al., 2014) Cuyo y Noreste argentino, particularmente Entre Ríos ( Segura et al., 2006; Guillén and Sánchez, 2007b; Liendo et al., 2012).



En Argentina, la primera liberación de *D. longicaudata* fue documentada en 1960, aunque las sucesivas liberaciones a campo no fueron evaluadas de manera adecuada (Ovruski and Fidalgo, 1994). En 1961 el Ministerio de Agricultura y Ganadería junto con el Instituto Nacional de Tecnología Agropecuaria (INTA) introdujeron a *D. longicaudata* desde México. Una cantidad de 100.000 individuos fue liberada en el noreste de las provincias de Misiones (Montecarlo) y Entre Ríos (Concordia) y en el noroeste de las provincias de Tucumán (San Miguel de Tucumán) y Jujuy (Calilegua). También algunos ejemplares fueron liberados en el centro del país, en Córdoba (Cruz del Eje y Yacanto) (Turica, 1968). Nuevas liberaciones fueron realizadas en Tucumán en 1977 y 1986 (Ovruski and Fidalgo, 1994). A principios de septiembre del año 1999 fue nuevamente introducido a fin de reanudar el empleo de enemigos naturales contra *C. capitata* y *A. fraterculus*. Desde ese momento hasta el presente, esta especie parasitoide está siendo criada a nivel semimasivo en la Planta Piloto de Procesos Industriales Microbiológicos (PROIMI) de San Miguel de Tucumán (provincia de Tucumán, Argentina) y a nivel experimental en el Instituto de Genética “Ewald A. Favret” (IGEAF), INTA Castelar (provincia de Buenos Aires, Argentina) (Ovruski et al., 2003a; Viscarret et al., 2006a, 2006b). En esa oportunidad el parasitoide se introdujo en nuestro país a través de un subsidio del Fondo Nacional de Ciencia y Tecnología (FONCyT, Argentina) desde el Complejo Bioindustrial MOSCAMED-MOSCAFRUT (Metapa de Domínguez, Chiapas, México) (SENASA, Expte. n° 14054/98). Recientemente se han llevado a cabo liberaciones del parasitoide en la provincia de San Juan y Mendoza, en el marco del ProCEM que incluye a *D. longicaudata* en adición a la TIE y el control por medios químicos (Suarez et al., 2014).

Relevamientos acerca de la presencia de parasitoides de moscas de la fruta han arrojado la presencia de especímenes de *D. longicaudata* en marzo de 2000, febrero y marzo de 2001 y abril de 2002 en la localidad de Montecarlo (Misiones, Argentina). Si bien se han recuperado pocos individuos, se puede decir que el parasitoide se ha establecido con éxito, al menos en una provincia argentina (alrededor de 1 % de parasitismo) (Schliserman et al., 2003b).

## 2.4. Genética de *D. longicaudata*

Particularmente para *D. longicaudata* se ha generado en los últimos años una vasta cantidad de información acerca de su biología, comportamiento, fisiología y se han comenzado a explorar aspectos de su genética (Kitthawee, 2000; Tejangkura et al., 2004; Paranhos et al., 2008; Stuhl et al., 2011; Julsirikul et al., 2014). Asimismo, se han hecho avances recientes en la caracterización taxonómica de la especie, definiéndola en realidad como un complejo de especies crípticas (Kitthawee and Dujardin, 2009). Este hecho es de importancia a la hora de

analizar su comportamiento a campo ya que podrían existir diferencias de *performance* que deben ser consideradas en las estrategias de control biológico implementadas en regiones con poblaciones naturales de este parasitoide.

Son de particular interés los conocimientos generados recientemente en genética y citogenética, que han permitido la caracterización de los estadios del desarrollo (Carabajal Paladino et al., 2010) y del cariotipo del parasitoide (Carabajal Paladino et al., 2013). Aún más recientemente se ha incluido a *D. longicaudata* entre las más de 60 especies de himenópteros que presentan determinación sexual por medio del sistema de alelos complementarios con al menos dos loci involucrados (Carabajal Paladino et al., 2015).

# Mecanismos de Determinación del Sexo

### 3. Determinación del sexo

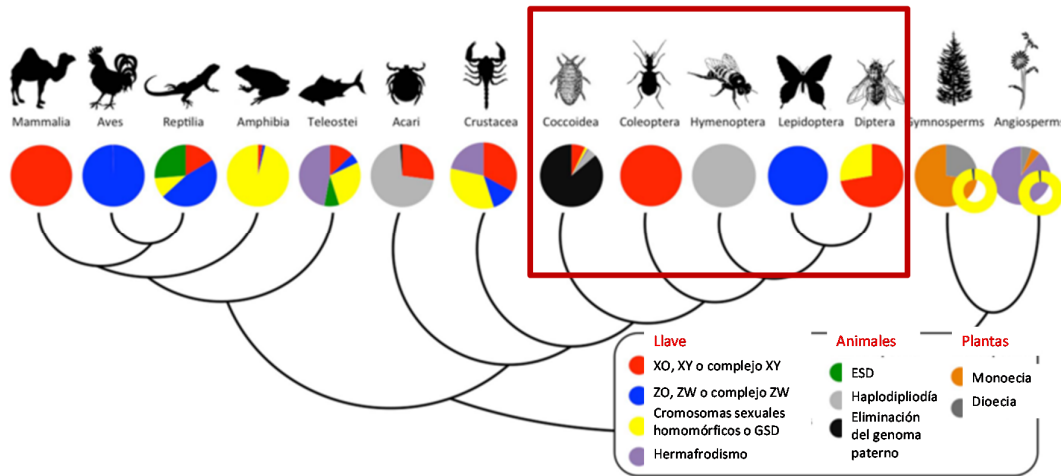
La determinación del sexo es una parte integral de la reproducción y forma parte de un proceso esencial para la evolución y enriquecimiento del genoma. Por estos motivos se trata de un tema que ha sido y es objeto de estudio y foco de interés para todas las especies que conforman el reino animal. La mayoría de los organismos eucariotas se reproducen sexualmente para producir una descendencia genéticamente diversa (Figura 3) (Bachtrog et al., 2014). Las diferencias fisiológicas, morfologías y comportamentales observadas entre machos y hembras son el resultado de la integración de dos procesos: la determinación del sexo y la diferenciación sexual (Normark, 2003).

La determinación del sexo comprende el conjunto de factores y mecanismos genéticos que definen el carácter sexo de un individuo (Lacadena, 1988). De este modo, la determinación del sexo se refiere al programa de desarrollo de un individuo que lo compromete hacia el camino femenino o masculino (Sánchez, 2008). Desde los insectos hasta los mamíferos aún existen numerosas incógnitas de los múltiples mecanismos empleados para llevar a cabo la determinación sexual. En el reino animal, a partir del estudio de varios organismos modelo, los diferentes datos obtenidos sugieren que la regulación molecular de dichos mecanismos se encuentra sujeta a una de las siguientes categorías:

- a. Acción ambiental sobre el embrión en un momento crucial del desarrollo: Principalmente se refiere a la variación de la temperatura en momentos determinados del desarrollo (períodos termosensibles) y se observa mayoritariamente en peces y reptiles (Vitková, 2007).
- b. Acción de la genética: Se considera así cuando el regulador central en la cascada de eventos que desencadenan la determinación del fenotipo sexual es al menos un gen específico. Este tipo de mecanismo es característico de algunas especies animales, incluyendo a los invertebrados y anfibios (Vitková, 2007).
- c. Presencia de cromosomas sexuales diferentes: El par sexual idéntico puede estar presente en ambos, machos (pájaros) y hembras (mamíferos), y el gen determinante puede ser bien conocido (como el SRY de los mamíferos) o sospechado.

Para esta tesis será de particular interés la determinación del sexo por acción de la genética, en donde uno o más genes son los responsables de la regulación que decidirá el destino de un cigoto independientemente de las condiciones ambientales. Estos genes se denominan de determinación del sexo y serán los responsables de la regulación de aquellos que se denominan

de diferenciación del sexo, encargados de dar origen a la formación de estructuras sexualmente dimórficas que caracterizan a la hembra y al macho adultos (Sánchez, 2008).



**Figura 3.** Diversidad de los sistemas de determinación sexual incluyendo representantes de las diferentes ramas (Bachtrog et al., 2014).

### 3.1. Determinación del sexo en insectos

Los mecanismos de determinación del sexo pueden ser muy variables entre especies cercanamente relacionadas a nivel filogenético. Aparentemente, el mismo mecanismo puede estar regulado por diferentes genes, sugiriendo un sistema de rápida evolución (Hasselmann et al., 2008; Gempe, T; Beye, 2011; Bopp et al., 2014). En insectos, las vías de determinación del sexo se componen de una cascada génica que es altamente variable en la parte superior, pero comparten un bloque conservado que incluye un regulador binario común, el gen *transformer* (*tra*), y al menos, un ejecutor común en la parte inferior, el gen *doblesexo* (*dsx*) (Bopp et al., 2014; Geuverink and Beukeboom, 2014). Una señal primaria inicia una de las dos rutas alternativas en la cascada de señalización que conducirá al “splicing” alternativo del ARNm del gen *doublesex* (*dsx*) que será sexo-específico y del gen *sex lethal* (*sxl*) que controla la diferenciación sexual (Bopp et al., 2014) ver detalles en Figura 4.

### 3.2. Determinación del sexo en Hymenoptera (antecedentes principales en *A. mellifera* y *N. vitripennis*)

En himenópteros, se han propuesto varios modelos que permiten describir las vías de determinación del sexo, sin embargo, estudios funcionales de los genes involucrados sólo se han llevado a cabo para algunas especies. El modelo más estudiado es el insecto social *Apis mellifera* L. (Hymenoptera: Apidae), la abeja melífera/doméstica, que presenta un sistema de determinación sexual denominado de alelos complementarios de un único locus (sl-CSD) (Beye et al., 2003). El desarrollo diferencial correspondiente a cada sexo bajo este sistema depende de la composición alélica de un único locus sexual. Los individuos heterocigotas para el locus de determinación sexual complementario (*csd*) se convierten en hembras y los hemicigotas o homocigotas se convierten en machos (Beye, 2004). La combinación heteroalélica de CSD activa el factor de “splicing” alternativo *transformer* (*tra*), denominado *feminizer* (*fem*) en el modelo de *A. mellifera*, que regula el “splicing” alternativo del pre-ARNm de *dsx* en los individuos que desarrollarán en hembras. En los homocigotas en cambio, no existe una variante funcional para la proteína TRA/FEM debido al “splicing” diferencial en aquellos individuos que desarrollaran en machos (Beye, 2004; Gempe, T; Beye, 2011). Otra especie que ha sido recientemente estudiada en profundidad es el parasitoide gregario *Nasonia vitripennis* Walker (Hymenoptera: Pteromalidae). *Nasonia spp.* posee un sistema de determinación del sexo descrito como el efecto materno de “imprinting” genómico (MEGISD). En este caso, la forma femenina específica de *Nvtra* (designación del gen *transformer* en *N. vitripennis*) depende de un círculo autorregulatorio, donde es necesario que sea superado un determinado umbral de concentración de ARNm materno provisto al huevo para que el individuo se convierta en una hembra (Verhulst et al., 2010).

### 3.3. Determinación del sexo en *D. longicaudata*

Los aportes a los conocimientos en genética y citogenética del complejo *D. longicaudata* han crecido en los últimos años. En relación a la citogenética se ha caracterizado el cariotipo como se ha mencionado anteriormente, y para esta especie haplodiploide los machos cuentan con 20 cromosomas y las hembras diploides con 40 (Carabajal Paladino et al., 2013).

Otro gran avance en el sentido de la genética de la determinación del sexo del parasitoide ha sido el estudio reciente en el que se propone el modelo de determinación del sexo por alelos complementarios (Carabajal Paladino et al., 2015). Se incluye de esta manera a *D. longicaudata* dentro de las más de 60 especies de himenópteros que presentan este sistema de determinación del sexo (Harpur et al., 2012). Se propone en este modelo que para *D. longicaudata* el sistema estaría compuesto por al menos dos loci (Carabajal Paladino et al., 2015).

# Tecnologías de secuenciación masiva: transcriptoma asociado al sexo

## 4. Pirosecuenciación 454: aplicaciones

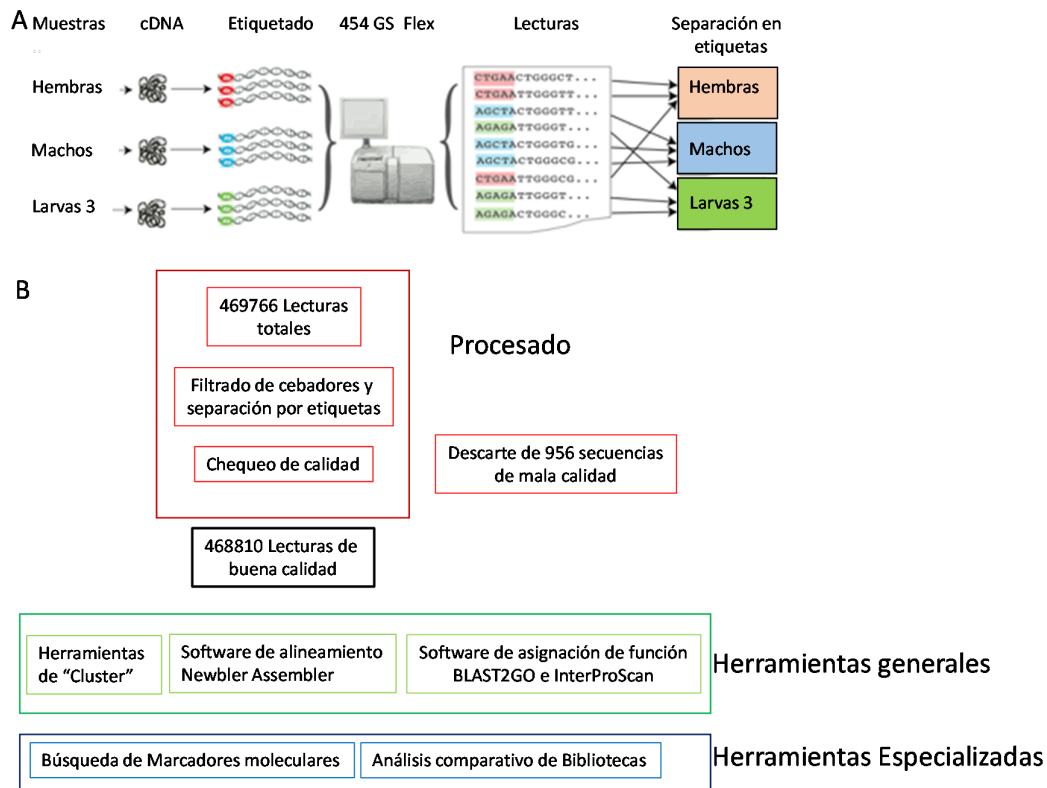
La reciente disponibilidad de métodos de secuenciación masiva (Roche 454, Solexa/ Illumina, etc.) a un bajo costo ha permitido la obtención rápida y a gran escala de datos acerca de genomas y transcriptomas de una inmensa variedad de especies no modelo, es decir, especies que no han sido ampliamente estudiadas (Vera et al., 2008). Entre estas se incluye una amplia variedad de artrópodos de la que existe poco o ningún conocimiento a nivel molecular, permitiendo de esta manera su aplicación como herramienta para la genómica funcional (Vera et al., 2008; Zou et al., 2008; Zagrobelny et al., 2009; Bai et al., 2010; Jaworski et al., 2010; Mittapalli et al., 2010; O'Neil et al., 2010; Olafson et al., 2010; Pauchet et al., 2010; Zhang et al., 2010).

En particular, la tecnología de pirosecuenciación 454 se basa en la detección de pirofosfato durante la síntesis de ADN y fue la primera alternativa al método tradicional de secuenciación de Sanger (1977). Es un método flexible que permite el procesamiento en paralelo de grandes cantidades de datos (Figura 4). Presenta además una amplia variedad de aplicaciones como identificación de marcadores moleculares (SNP y SSR), cuantificación de transcritos a partir de bibliotecas de ADN copia (ADNc) y secuenciación de genomas completos. Existen también una serie de desventajas como la dificultad para determinar el número de nucleótidos incorporados en regiones homopoliméricas que contienen más de 5 o 6 nucleótidos o la longitud de los fragmentos secuenciados. Si bien esta última ha mejorado mucho en los últimos años se espera que se logren superar estas dificultades.

Esta metodología consiste de la unión de fragmentos de hebras únicas de ADN a una esfera (una hebra por esfera), por medio de un adaptador. En el caso de tratarse de una muestra de ARN total eucariota, en primer lugar la misma se enriquece en ARNm a través de columnas poli(T) para luego sintetizar las bibliotecas de ADNc correspondientes. Esas hebras se fragmentan en secuencias de entre 500 y 800pb que serán las que se unirán a las esferas. En este punto, se adiciona la mezcla de enzimas necesarias para la amplificación y secuenciación: ADN polimerasa, sulfúrilasa, ATP luciferasa y apirasa, y los sustratos APS y luciferina. La amplificación se llevará a cabo en una emulsión de agua y aceite donde cada esfera quedará en un microambiente aislado. Al final de los ciclos de emPCR (PCR en emulsión), se desnaturalizan las hebras y cada esfera queda cargada de amplicones de una secuencia repetida muchas veces. La secuenciación se lleva a cabo sobre una microplaca "Pico TiterPlate" de 400000 pocillos de 44 micras de diámetro. En cada uno de estos pocillos se depositará una única esfera con amplicones y se colocará la microplaca en el secuenciador. En cada uno de los pocillos, el secuenciador repetirá ciclos que consistirán en vertido de un tipo de nucleótido único (A, T, G o C) seguido de un lavado y así



sucesivamente hasta completar la síntesis de las secuencias complementarias a las secuencias molde. Cada vez que la polimerasa introduzca una base, se producirá la liberación de pirofosfato, el cual será convertido en ATP a través de la sulfúrilasa en presencia de APS. El ATP generado, media en conjunto con la enzima luciferasa, la conversión a oxiluciferina, generando una emisión de luz. La degradación de dNTPs no incorporados y de ATP se produce por acción de la enzima apirasa. La luz emitida es detectada por una cámara CCD e interpretada por un software que construirá un pirograma con la secuencia de cada pocillo. El pirograma reflejará la cantidad de nucleótidos incorporados con una intensidad de luz emitida proporcional (Papanicolaou et al., 2009). Si la polimerasa no incluye nucleótidos en un ciclo, el pirograma dejará huecos hasta que se incluya el nucleótido correcto.



**Figura 4.** Esquema de trabajo y procesamiento de datos. Se muestra el procesamiento general de la muestra en A y en B el análisis de secuencias obtenidas.

#### 4.1. Aplicaciones de la pirosecuenciación: descubrimiento de genes y marcadores

La principal aplicación de las tecnologías de secuenciación masiva consiste en proveer grandes cantidades de información en tiempos cortos y a costos relativamente bajos. Muchas especies de las que se estudian características fisiológicas o ecológicas no cuentan con datos acerca de su genómica o transcriptómica. Para la obtención de genomas y/o transcriptomas de especies no modelo, se recurre tradicionalmente a tecnologías tales como construcción de bibliotecas de ADNc, rondas repetidas de normalización/sustracción y secuenciación por el método de Sanger, y habitualmente se contruyen “microarrays” de ADNc (Beldade et al., 2006; Mita et al., 2003; Papanicolaou et al., 2005; Paschall et al., 2004; Rudd, 2003; Whitfield et al., 2002). Con el surgimiento de los métodos de NGS se ha logrado complementar y potencialmente superar estas tecnologías. Asimismo, se reducen costos, son metodologías menos laboriosas (Hudson, 2008), se reducen los errores por mal manejo de clones y permite recuperar transcriptos poco frecuentes o inestables cuando se clona en sistemas bacterianos (Weber et al., 2007). Es decir, estas tecnologías permiten la obtención, el análisis y caracterización de genomas y transcriptomas de especies no modelo de forma rápida y precisa (Vera et al., 2008). Como parte del análisis del transcriptoma es posible identificar grupos de genes o transcriptos de interés (Zagrobelny et al., 2009; Smith et al., 2011; Yúfera et al., 2011) y conocer datos de expresión diferencial entre muestras analizadas a gran escala (Delano-Frier et al., 2011; Eo et al., 2011; Roulin et al., 2014).

Los microsatélites o Secuencias de Repetición Simple (SSR de su sigla en inglés “Simple Sequence Repeats”) son encontrados en toda la extensión del genoma eucariota (Toth et al 2000) y se localizan tanto en regiones codificantes como no codificantes (Powell 1996; Toth 2000). Los SSR han sido uno de los sistemas de marcadores moleculares más útiles por su herencia por co-dominancia, abundancia en los genomas, alta reproducibilidad y su habitual uso en investigación genética (Kumar et al., 2014). El desarrollo de SSR a partir de secuencias expresadas, por medio de transcripción reversa (de sus siglas en inglés “reverse transcription” RT) de ARN, ha surgido como una nueva estrategia para la rápida obtención de SSR de un modo confiable y relativamente económico (Silva et al., 2013; Wei et al., 2011; Zhang et al., 2012; Zheng et al., 2013). Debido a su existencia en regiones del genoma que son transcripcionalmente activas, los marcadores producidos EST-SSR (SSR provenientes de Secuencias Expresadas, del inglés “Expressed Sequence Tags-SSR”) pueden llevar al desarrollo de mapas génicos que colaboran con la identificación funcional de genes e incrementar la eficiencia de la selección asistida por marcadores (Gupta et al., 2004; Varshney et al., 2005). Adicionalmente, los SSR provenientes de secuencias expresadas muestran un mayor nivel de transferibilidad entre especies relacionadas filogenéticamente que los SSR provenientes de secuencias genómicas (Saha et al., 2004); de hecho han podido desarrollarse con éxito marcadores a partir de secuencias expresadas para

muchas especies (Aggarwall et al., 2007; Choudhary et al., 2009; Feng et al., 2009; Qiu et al., 2009).

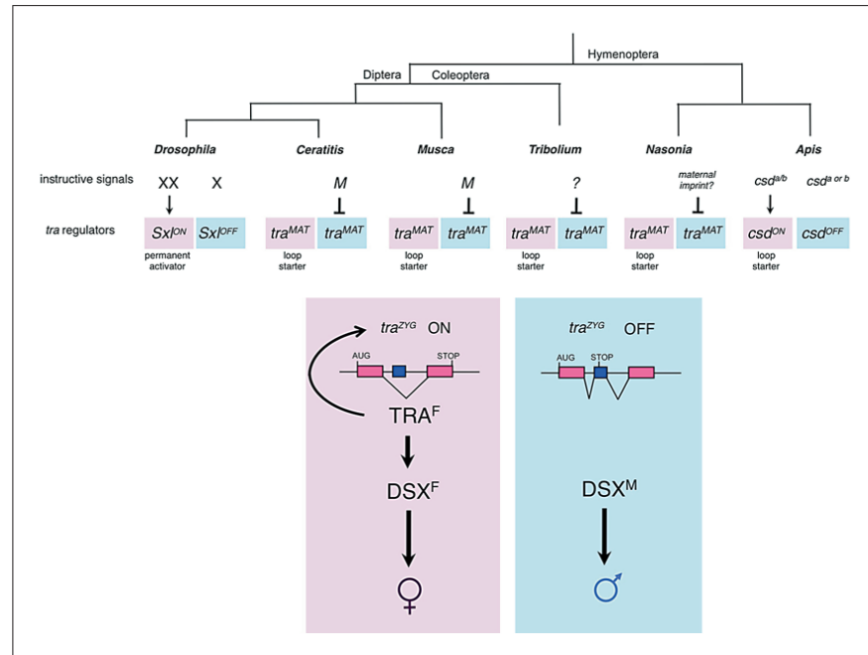
# RESULTADOS

# Extracción de ARN y Búsqueda de Genes Candidatos

# 1. Identificación de genes involucrados en la determinación del sexo por homología

Para iniciar el estudio de las bases moleculares de la determinación del sexo en *D. longicaudata* se utilizó la información disponible de los organismos tradicionalmente estudiados como modelo dentro de los himenópteros: *A. mellifera* y *N. vitripennis*. A partir de estos dos modelos se seleccionaron tres genes que se encuentran caracterizados y se los vincula a los mecanismos de determinación del sexo en esas especies: *complementary sex determiner* (*csd*); *feminizer* (*fem*) y *doublesex* (*dsx*) (Figura 5).

En trabajos preliminares del grupo (Lanzavecchia et al., 2008) realizados a partir ADN total del parasitoide y utilizando cebadores ya diseñados para *A. mellifera*, si intentó identificar secuencias homologas al gen *csd* en el genoma del parasitoide. Se observaron dos productos de amplificación que no presentaron homología en las bases de datos de secuencias nucleotídicas.



**Figura 5.** Sistemas de determinación del sexo en Hymenoptera. Alternativas regulatorias para el sistema de determinación del sexo en individuos haplodiploides que poseen Sistema de Determinación del Sexo por alelos Complementarios o por medio de “Imprinting” materno (Bopp et al., 2014).

## 1.1. Puesta a punto de la extracción de ARN

En base a la identificación de dos fragmentos obtenidos previamente por medio de PCR inespecífica sobre ADN genómico (Lanzavecchia et al., 2008), se decidió evaluar si estas regiones presentes en el genoma del parasitoide se encontraban expresadas. Por este motivo, fue necesaria la puesta a punto de la extracción de ARN para *D. longicaudata*. Para ello, se ajustaron las condiciones descritas por el fabricante para la extracción de ARN a partir de Trizol (Invitrogen) a fin de obtener resultados óptimos y ARN de alta calidad para evaluar la transcripción de estas regiones, en primera instancia por PCR de punto final y a *posteriori* por RT-qPCR.

### 1.1.1. Volumen de TRIzol en relación a la masa de tejido de partida

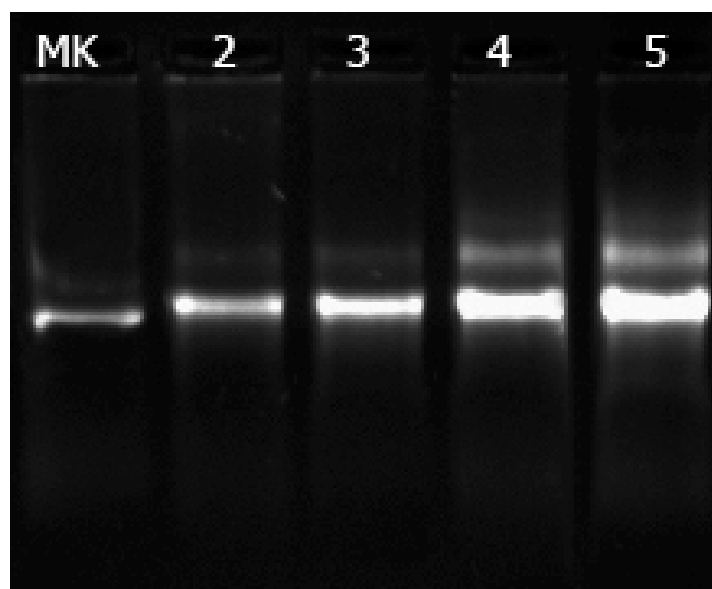
La extracción de ARN suele ser un proceso complicado debido a la elevada labilidad del material en sí mismo y a la presencia de ARNasas, ya sea dentro de la muestra o en el ambiente. En la actualidad la cantidad de técnicas en biología molecular que utilizan como material de partida ARN o ADNc es cada vez mayor y en el caso de los insectos es escasa la información publicada en relación a protocolos de extracción de ARN (Ridgeway and Timm, 2014). Por estas razones es de vital importancia analizar la calidad del material obtenido para asegurar resultados confiables. Una muestra que se encuentre intacta debería, por definición, contar con la presencia clara de las dos bandas características del ARN ribosomal (ARNr) eucariótico correspondientes a las subunidades 28S y 18S. Asumiendo que la integridad del ARNr refleja la integridad de otras fracciones de ARN, como el ARNm que es de interés en este caso, es necesario detectar la presencia de ambas al llevar a cabo el análisis por medio de geles de agarosa teñidos con Bromuro de Etidio (Winnebeck et al., 2010). Cabe destacar que, para una correcta interpretación de los datos de calidad de material, en el caso de los insectos a diferencia de otros eucariotas, los perfiles de ARNr suelen presentar una única banda clara. Esta banda correspondiente a las dos fracciones 18S y 28S, se genera porque la fracción 28S del ARNr de insectos es altamente termolábil. No debe asumirse que la muestra se encuentra degradada al observar una banda muy intensa y una leve sombra de mayor tamaño (Winnebeck et al., 2010). Asimismo, debe observarse la presencia de contaminación con ADN que se evidenciaría como una banda de mayor tamaño o inclusive como fluorescencia en el pocillo de sembrado de la muestra dependiendo de las condiciones de corrida. Comúnmente la contaminación con ADN genómico no llega a evidenciarse en geles de agarosa teñidos con Bromuro de Etidio o se visualiza de manera tenue, por lo que es necesario llevar a cabo controles (uso del ARN como molde en las reacciones de qPCR) no solo en el paso de extracción de ARN (Evrogen, 2015). Se

retomará este punto más adelante al momento de utilizar el material para las posteriores aplicaciones.

Para comenzar con la puesta a punto y establecimiento de un protocolo de extracción aplicable a *D. longicaudata* y sus particularidades biológicas, se probaron las extracciones de ARN a partir de 5, 10, 15 y 30 individuos adultos. El mejor resultado de integridad, cantidad y calidad de material se obtuvo con la utilización de grupos de 15 individuos adultos (alrededor de 45mg de tejido). Cuando se analizó el ARN obtenido mediante gel de agarosa 1%p/v (Biodynamics), se pudieron observar las dos bandas características correspondientes a los ARNr de mayor abundancia en eucariotas 18S y 28S, con el característico aspecto difuso de la banda de ARN correspondiente a la subunidad ribosomal mayor. Asimismo, no se observó presencia de material degradado que se evidencia como bandas inespecíficas o una banda difusa en el frente de corrida del gel. Por estas razones puede decirse que el material obtenido posee una buena integridad. En cuanto a la evidencia de contaminación con ADN genómico, en principio no se observa en el gel de agarosa. A pesar de que esto es un buen indicador, el análisis en este caso se retomará más adelante en la sección 1.1.2.

Los resultados de calidad del material obtenidos por medio de las relaciones de absorbancia (NanoDrop ND-1000) respaldaron las observaciones en gel, en todos los casos se obtuvieron valores cercanos a los óptimos (alrededor de 2, según especificaciones de NanoDrop) para los parámetros que evalúan contaminación de la muestra por presencia de solventes orgánicos (relación de absorbancia 260/230), así como contaminación por presencia de proteínas (relación de absorbancia 260/280) (Tabla1). Al analizar las concentraciones de ARN obtenidas para los diferentes grupos se evidenció la primera gran diferencia, al utilizar números de individuos menores a 15 no se obtenía la cantidad mínima de 1 ug de ARN requerida para llevar a cabo la transcripción reversa (RT) según el protocolo del fabricante (Promega). De esta manera, no era posible obtener buenos resultados de expresión o las muestras se encontraban en el límite inferior de concentración y luego no se obtenían buenos resultados de amplificación por PCR (Tabla1, Figura 6). Cuando se partió de 30 individuos en un mismo grupo, la calidad del ARN obtenida mejoró sustancialmente al utilizar el doble del TRIzol recomendado, reflejo de esto son los índices de calidad en función de la contaminación con proteínas y solventes orgánicos que se acercan al valor óptimo. Asimismo se observaron mejores resultados en las amplificaciones posteriores. Este primer protocolo, partiendo de 15 individuos, se utilizó en adelante para realizar extracciones de ARN cuyo objetivo fuera realizar PCR, ya sea de punto final como qPCR.





**Figura 6.** Extracción de ARN para cantidades crecientes de individuos por grupo. Calle1: marcador de peso molecular 1000pb y 750pb; calles 2 a 5: 5, 10, 15 y 30 individuos respectivamente.

**Tabla1.** Parámetros de calidad para el ARN obtenido.

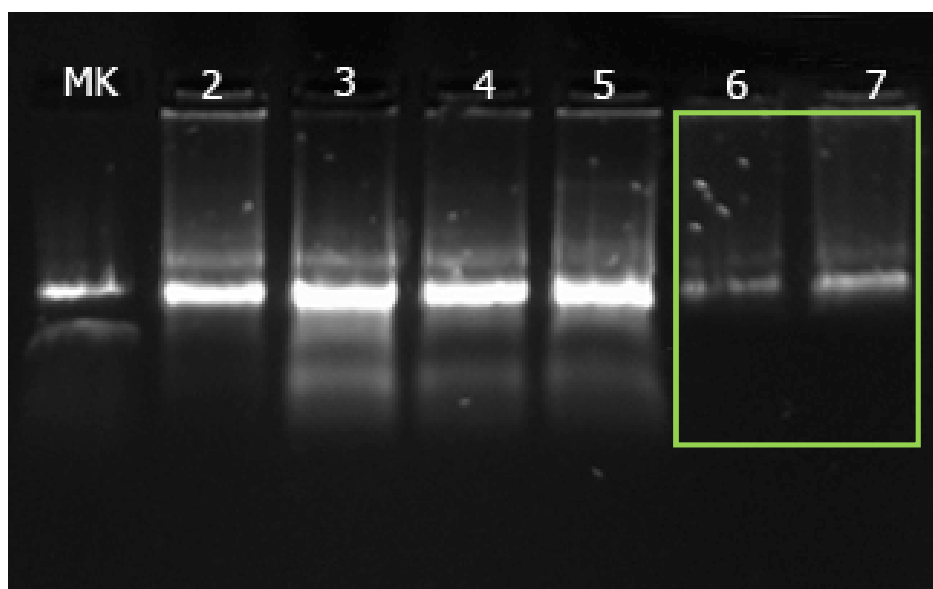
Muestra	ug/ul	260/280	260/230
5 adultos	520	1,90	1,75
10 adultos	1300	1,80	1,83
15 adultos	5700	2,10	1,92
30 adultos	6200	2,20	2,35

### 1.1.2. Con y sin uso de ADNasaI

Algunas aplicaciones para las cuales es necesario contar con ARN como material de partida requieren que este sea de alta calidad y fundamentalmente que se evite la contaminación por presencia de ADN genómico. Una de estas aplicaciones de especial sensibilidad es la RT-qPCR, por lo que el ARN que será usado posteriormente para llevar a cabo tanto reacciones de transcripción reversa (RT) a ADNc, como las reacciones de amplificación en tiempo real (RT-qPCR), no debe presentar concentraciones detectables de material genómico. Es posible, como se mencionó anteriormente, que la presencia de ADN no se detecte en gel de agarosa teñido con Bromuro de Etidio. No obstante, la sensibilidad de las reacciones de RT-qPCR puede detectar

concentraciones muy bajas de ácido nucleico, por lo cual una mínima contaminación será detectable por qPCR alterando la fidelidad de los resultados obtenidos. Por este motivo se trataron las muestras con una enzima que degrada el ADN (ADNasa I, DNase I kit, Promega) a fin de evaluar la calidad de ARN obtenido a partir de los resultados de amplificación por PCR y RT-qPCR comparando muestras tratadas y no tratadas con ADNasa I (Figura 7, Tabla 2).

Las muestras sin tratamiento (calles 2-5 Figura 7) presentan bandas intensas, aunque en algunos casos se observa ácido nucleico degradado, que también se evidencia como una banda difusa en el frente de corrida e incluso una banda muy tenue de mayor tamaño a las dos bandas características para ARNr que corresponde al ADN genómico. Las muestras de ARN tratadas con ADNasa I (calles 6 y 7 en Figura 7), evidenciaron una gran pérdida de material genético (bandas tenues comparadas con las muestras sin tratamiento con ADNasa I). No se observa el fondo claro como en algunas de las muestras sin tratar y no queda rastro de la banda tenue de mayor tamaño correspondiente a ADN genómico. A fin de obtener la cantidad de ARN adecuada para los ensayos de RT y PCR, puede evitarse el uso de la enzima si se controla por medio de electroforesis la calidad del material obtenido y se evalúan en cada caso los resultados de RT-qPCR ya que en este sistema en particular no se evidencia su presencia por medio de PCR de punto final. Para controlar la presencia de ADN genómico que pudiera alterar las mediciones de qPCR, se incluye en el ensayo una muestra que utiliza como molde ARN sin previa realización de RT. La finalidad de incluir esta muestra radica en que en dicha reacción no debería existir amplificación. Las altas temperaturas degradan el ARN y si se evidencia amplificación indica presencia de ADN genómico y dicha muestra debería ser descartada. En ninguno de los casos se evidenció amplificación en estos controles por medio de reacciones de RT-qPCR.



**Figura 7.** Corrida electroforética e gel de agarosa 1% p/v de ARN total de individuos de *D. longicaudata*. En la primera calle se sembró marcador molecular (1000pb y 750pb); calles 2 y 3 muestras de ARN total de adultos de *D. longicaudata* sin tratamiento con DNAsa; calle 4 y 5 muestras de ARN total de adultos de *A. mellifera* sin tratamiento con DNAsa (control); calles 6 y 7 ARN total de *D. longicaudata* y *A. mellifera* tratados con ADNasa I, respectivamente.

**Tabla 2.** Calidad de extracción de ARN con y sin el uso de ADNasa I. Medidas de calidad de ARN obtenidas espectrofotométricamente (Nanodrop1000).

Muestra	ug/ul	260/280	260/230
<i>Apis mellifera</i> (con DNAsa)	1351,3	1,9	2,2
<i>Apis mellifera</i> (sin DNAsa)	4856,3	2,2	1,9
<i>D. longicaudata</i> (con DNAsa)	2344,2	2,2	2,1
<i>D. longicaudata</i> (sin DNAsa)	3222,4	2,1	2,0

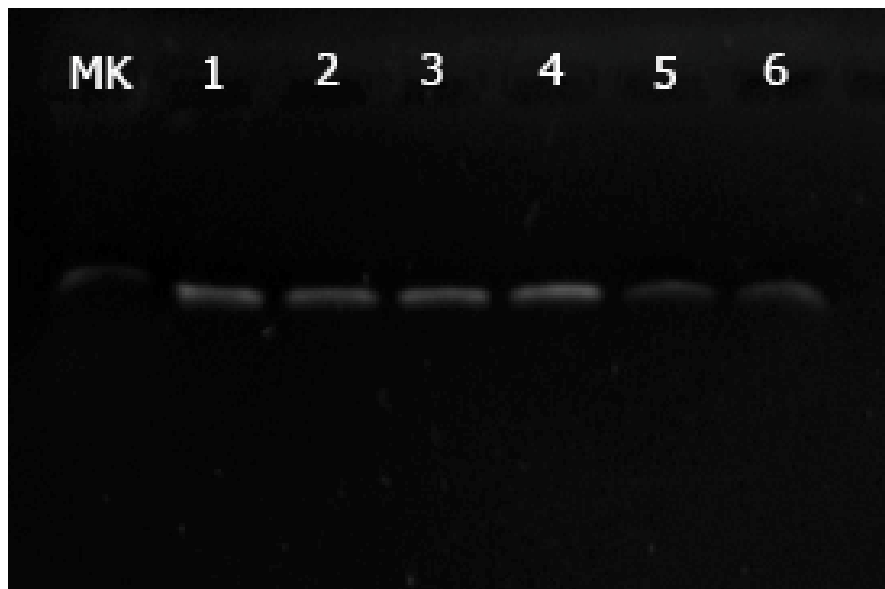
### 1.1.3. Puesta a punto de la Transcripción Reversa

Para obtener ADNc que permitiera la correcta amplificación de los fragmentos deseados por medio de la técnica de PCR de punto final y RT-qPCR, fueron necesarios ajustes al protocolo propuesto por el fabricante (kitImProm-II Reverse Transcriptase, Promega). Se analizaron las condiciones en las que se llevaría a cabo la RT a fin de obtener un protocolo que generase ADNc

de integridad suficiente para lograr una amplificación eficiente tanto por PCR como por qPCR de los fragmentos deseados.

El primer factor a definir fue qué tipo de cebador se utilizaría en las reacciones, para lo que se analizaron dos tipos: cebadores “random”, secuencias nucleotídicas cortas aleatorias que hibridizan a lo largo de todas las cadenas de ARN y cebadores “Oligo(dT)”, estos últimos contienen en su secuencia repeticiones de T a fin de optimizar la RT a partir de ARNm eucariota exclusivamente. Utilizando diluciones de ADNc obtenido a partir de los diferentes cebadores (“random” y “Oligo(dT)”) como molde para reacciones, tanto de PCR como de qPCR, pudo observarse que los resultados no se veían afectados por el uso de cualquiera de los dos cebadores (Figura 8). Con la intención de evaluar regiones expresadas en el parasitoide, se decidió la utilización de cebadores Oligo(dT).

En relación a los cebadores también fue necesario ajustar la concentración final de los mismos en la mezcla de reacción ya que con la concentración recomendada por el fabricante del kit de RT no se lograban buenos resultados de PCR (bandas débiles o difusas).



**Figura 8.** Productos de amplificación por RT-qPCR de  $\beta$ -actina con cebadores específicos para *D. longicaudata* a partir de ADNc de *D. longicaudata* obtenido con cebadores Random (calles 2 y 3) y cebadores OligodT (calles 4 y 5). En las calles 6 y 7 se observan los controles con *A. mellifera*. Marcador de peso molecular de 200pb (calle1).

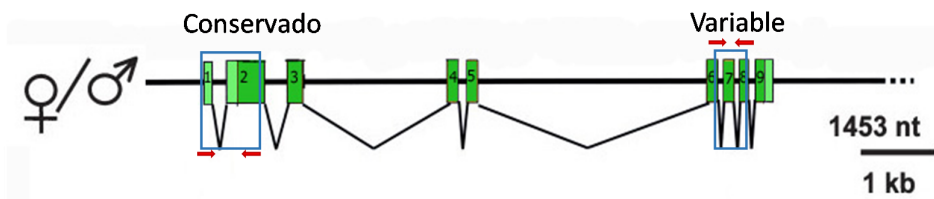
## 1.2. PCRs inespecíficas de los Genes de Interés: *csd* y *dsx*.

Siguiendo la línea de lo planteado en la sección 1.1 y para comenzar a explorar los posibles reguladores y efectores presentes en las vías de determinación del sexo de *D. longicaudata*, se analizaron dos de los genes presentes en el módulo central conservado de la cascada: el regulador *csd* descrito en las vías regulatorias de determinación sexual en *A. mellifera* y el efector *dsx*, descrito en la mayoría de las especies sexuales.

### 1.2.1. Gen *csd*

A partir de cebadores diseñados previamente en el grupo de trabajo sobre ADN de *D. longicaudata* en base a homología de secuencias de nucleótidos con el locus *csd* de *A. mellifera*, se logró amplificar por medio de ciclos de PCR fragmentos de ADNc.

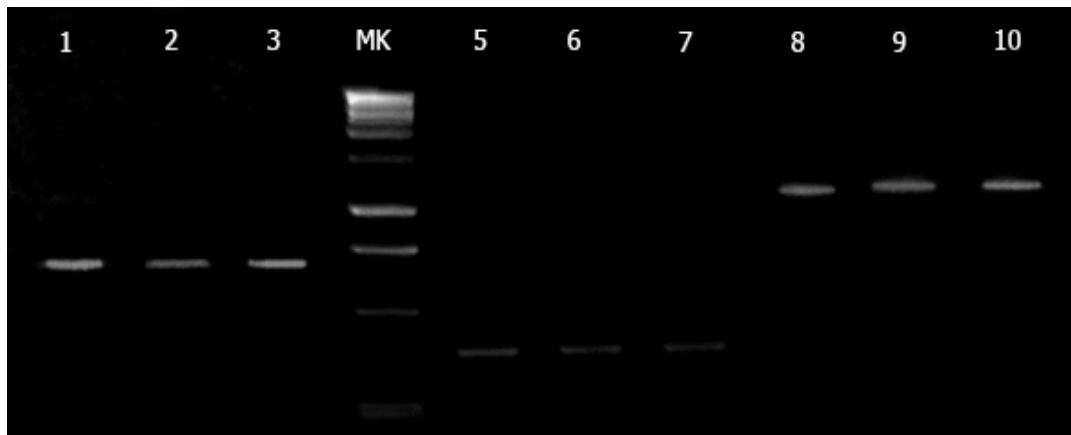
#### Esquema del gen *csd*



**Figura 9.** Ubicación de los cebadores específicos para *csd* sobre el esquema del gen para *A. mellifera* (Beye 2003). Se representa la ubicación de los nueve exones en cajas verdes y de intrones en líneas negras. Los cebadores utilizados se indican con flechas rojas.

Los resultados de las diferentes combinaciones de cebadores (Figura 9) arrojaron los mismos tamaños de fragmento amplificado para ambos sexos (Figura 10). Comparando con los fragmentos obtenidos sobre ADN genómico (Lanzavecchia et al., 2008), los fragmentos obtenidos sobre ADNc presentaron el mismo tamaño. Para la denominada Región Conservada (calles 1-3) el producto de PCR fue de 700bp tanto en machos como en hembras. No se observaron diferencias de tamaño, ni entre sexos, ni con los fragmentos de ADN genómico. Al

secuenciar dicho producto de amplificación se encontró que la secuencia era coincidente con la previamente obtenida para los fragmentos de ADN genómico (Anexo 1). En el caso de la denominada Región Variable (calles 5-7), el producto de PCR de 400bp con cebadores específicos de *D. longicaudata* tampoco difirió entre machos y hembras. Nuevamente las secuencias obtenidas pudieron alinearse perfectamente con las obtenidas previamente para los fragmentos de ADN genómico (Anexo 1). Finalmente, observando el esquema del gen construido para *A. mellifera*, se intentó probar que estos fragmentos obtenidos para *D. longicaudata* formaban parte de un mismo transcripto, correspondiente a dos regiones de un mismo gen. Con este objetivo, se pretendió amplificar la región central flanqueada por estas regiones previamente descritas. En este caso se obtuvo un producto de 1500bp (calles 8-10). Se observó el mismo producto de PCR para todos los ADNc analizados, no pudiendo ser secuenciado en ninguna de las oportunidades.

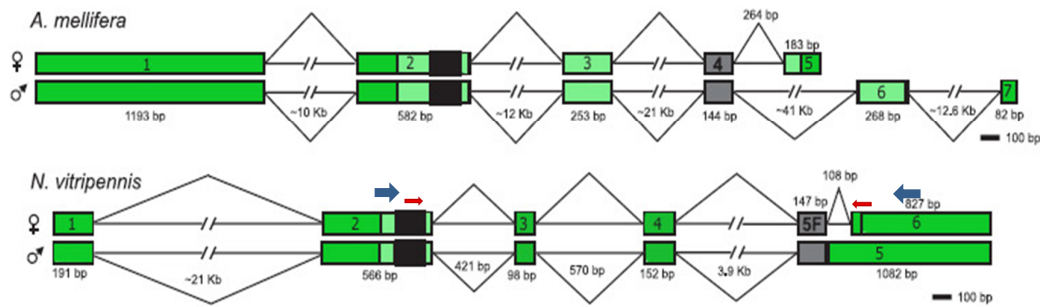


**Figura 10.** Productos de PCR para los cebadores *csd*. Las calles 1, 2 y 3 muestran el fragmento denominado conservado de 700pb para ADNc de hembras, machos y DNA respectivamente del parasitoide. Las calles 5 a 6 muestran el fragmento denominado variable de 400pb para ADNc de hembras, machos y ADN de *D. longicaudata* respectivamente. Las calles 8 y 9 muestran el fragmento de alrededor de 1500bp correspondiente a ADNc de hembras y machos de *D. longicaudata*. Se utilizó como marcador de peso molecular 1Kb ADN Ladder (Promega), calle 4. Control +: ADNc *A. mellifera*, calle 10.

### 1.2.2. Gen *dsx*

Se estudió la presencia del gen efector *dsx* mediante el diseño de cebadores inespecíficos sobre las regiones expresadas a partir de alineamientos que presentaron mayor homología de secuencia, tanto a nivel de proteína como de nucleótidos (Anexo 2). En la figura 11 se muestra la ubicación de los cebadores diseñados sobre el esquema del ARNm proveniente del gen *dsx* en las dos especies modelo dentro del orden Hymenoptera. Utilizando estos cebadores se intentó amplificar el efector *dsx* para *D. longicaudata* sobre ADNc y ADN genómico.

### Esquema del gen *dsx* para organismos modelo



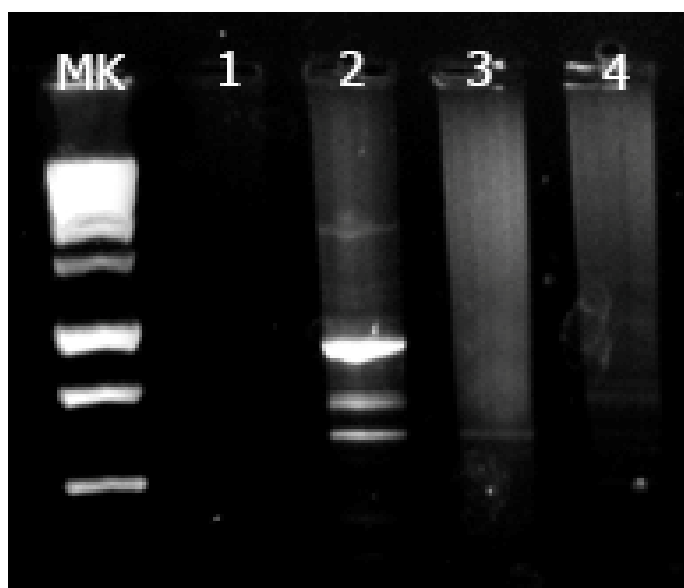
**Figura 11.** Esquema del gen *dsx* para ambos sexos en las dos especies modelo del orden Hymenoptera más ampliamente estudiadas (Oliveira et al., 2009). Las cajas verdes representan la ubicación de los exones y los bloques negros y grises representan el Dominio DM (unión al ADN) y el Dominio DSX (dimerización) respectivamente, las líneas negras identifican los intrones. Los cebadores diseñados se muestran en flechas rojas y azules.

**Tabla 3.** Cebadores degenerados diseñados para amplificar las diferentes regiones *dsx* en *D. longicaudata*. Para ver la ubicación de cada par de cebadores ver el Anexo 2.

Primers	secuencia	Tm	Posición	Exón
Dsx6_fw	tggtcaccatttgccacagac	60,1	1324-1346	5
Dsx_rv	tcctctttggattgacctgttct	56,1	402-424	2
dsxF2	ncgatntcagaatnacggca	61	599-619	2
dsxFC	ntgtcagaatnacgnaaganggtcca	70	602-628	2
dsxFA	aganagganagantacaccacactg	59	160-185	1
dsxFB	aattgacaaactngtttacnaaa	54	323-342	2
dsxF19	cggtgacatgcgtagtttgagcaataa	69	850-878	4
dsxR2	atgtgnatactcacgnccaana	60	677-699	6
dsxRV	agcgatcgagagtcgngnctggg	72	524-546	2
dsxR37	agtatggatcagtcggatgacatggtg	68	494-518	2
dsxR19	gacagcaatagcagcgacggtgacat	71	850-878	4

dsxR10	gncaatcgntaactntgatgtacgta	61	1105-1130	5-6
dsxR6	aatggctgtgaatnctngtacctga	64	1430-1453	6

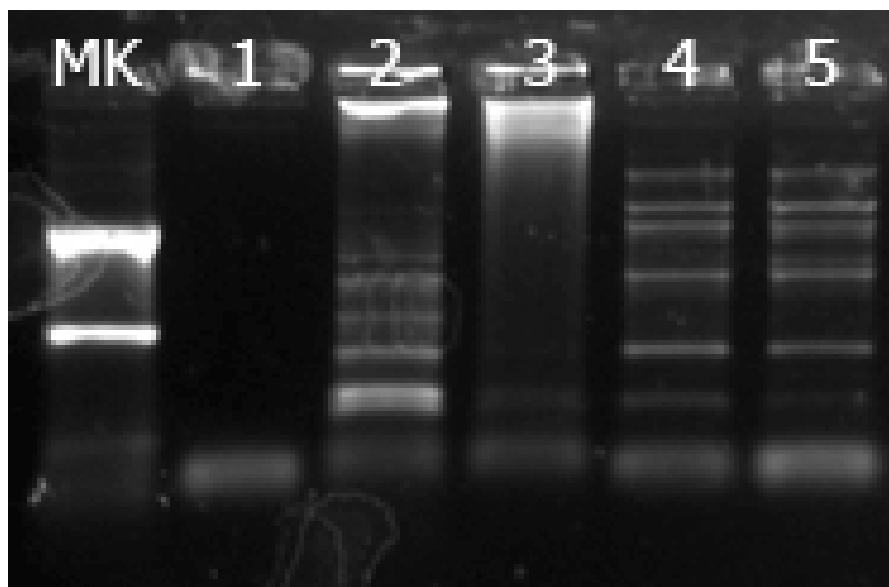
A partir de diversas combinaciones entre los cebadores diseñados, se lograron dos que permitieron la amplificación diferencial entre sexos. En la Figura 11 puede observarse las dos combinaciones de cebadores ubicados sobre el modelo de gen construido para las especies modelo del orden Hymenoptera. Para la primera combinación que generó productos de amplificación diferenciales, indicada con flechas rojas (Figura 11), se muestra la amplificación por PCR de punto final para muestras de ADNc y ADN genómico de machos y hembras adultas del parasitoide en la Figura 12. Se distinguen principalmente tres bandas, de 1000bp, 750bp y alrededor de 600bp respectivamente en la calle correspondiente a la muestra de hembras adultas. Tanto en machos como en la muestra de ADN no se vio producto de amplificación. Al intentar purificar y secuenciar estas muestras no se consiguieron en ninguna oportunidad secuencias de buena calidad que permitieran un alineamiento.



**Figura 12.** Productos de PCR para el primer par de cebadores *dsx* que presentó diferencias asociadas al sexo. Se utilizó como marcador de peso molecular 1Kb ADN Ladder (Promega). En la calle 1 se muestra el control negativo de reacción; en la calle 2 se observa ADNc de hembras adultas de *D. longicaudata*; en la calle3 se muestra ADNc de machos adultos y en la calle 4 ADN genómico.



En el segundo caso, se pueden observar los cebadores utilizados ubicados sobre el esquema del gen *dsx* para especies modelo del orden Hymenoptera como un par de flechas azules. Al igual que en el par anterior, también con este par de cebadores, se logran varias bandas de diferentes tamaños en la muestra correspondiente a ADNc provenientes de hembras adultas de *D. longicaudata* que no son observables en la muestra de ADNc de machos (Figura 13). A diferencia del par anterior, en este caso se logró también observar productos de amplificación en las muestras correspondientes a ADN genómico para ambos sexos. El patrón de amplificación fue idéntico en ambas muestras de ADN genómico. En el caso de las muestras correspondientes a ADNc, se observaron tres bandas de mayor tamaño en la muestra de machos (alrededor de 1200pb, 1000pb y 950pb) que las observadas en la muestra de ADNc de hembras. Continuando con la comparación entre los patrones de amplificación obtenidos en ADN genómico y en ADNc, se observan en ambas muestras bandas de 700bp y 350bp. El patrón de amplificación que se observa en hembras también posee una banda específica clara de alrededor de 450bp. También en este caso se purificaron las bandas y se intentó secuenciarlas, obteniendo secuencias de calidad muy pobre y poco fiables.



**Figura 13.** Productos de PCR para el segundo par de cebadores *dsx* que presentó diferencias asociadas al sexo. En la primera calle se observa el MK: 950-400bp. En la calle 1 se encuentra el control negativo de reacción; calles 2 y 3 ADNc de hembras y machos adultos de *D. longicaudata* respectivamente. Finalmente en las calles 4 y 5 se pueden observar los productos de amplificación correspondientes a ADN genómico de hembras y machos adultos de *D. longicaudata*.

Los resultados obtenidos hasta aquí evidencian que esta primer estrategia que aborda el estudio de identificación de genes involucrados en la cascada de determinación sexual en *D. longicaudata* por homología de secuencia en especies taxonómicamente relacionadas no resultó favorable para la obtención de secuencias específicas del parasitoide asociadas a esta vía metabólica. Por este motivo se decidió cambiar de rumbo de la investigación haciendo uso de las nuevas tecnologías de secuenciación masiva que han aportado considerablemente al estudio de sistemas no modelo permitiendo una rápida caracterización de los mismos.

# Análisis del Transcriptoma de *D.* *longicaudata*

## 2. Transcriptómica de *D. longicaudata*: Identificación de genes por Pirosecuenciación 454

Se utilizó la técnica de pirosecuenciación 454 con el objetivo de obtener datos de secuencias específicas de *D. longicaudata*, explorar el transcriptoma y detectar variantes de secuencia o en niveles de expresión de genes entre sexos. De esta manera se propone iniciar la caracterización de los mecanismos moleculares de determinación del sexo en el parasitoide.

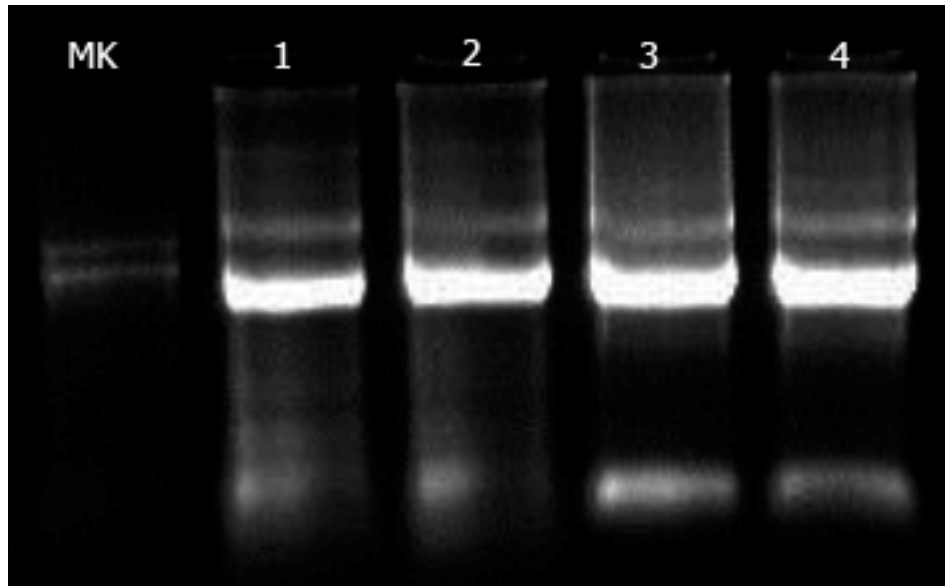
Para el análisis de los datos de transcriptómica, se desarrollaron “pipelines” para el ensamblado y análisis de los datos crudos generados por medio de la pirosecuenciación 454, de esa manera se logró el ensamblado *de novo* del transcriptoma así como la anotación y curado de la base de datos generada para *D. longicaudata* (Parra et al., 2007)

## 2.1. Caracterización del transcriptoma de *D. longicaudata*

### 2.1.1. Obtención de las muestras de ARN

La pirosecuenciación 454 de ADNc requiere muestras de ARN de alta calidad y concentración que no presenten contaminación genómica. A estos fines fue necesario ajustar las condiciones de extracción de manera de obtener material de la calidad e integridad requeridas. Los mejores resultados obtenidos se muestran en la Figura 14 y Tabla 4.

Para la obtención de datos a fin de explorar el transcriptoma de *D. longicaudata* se diseñó el experimento incluyendo muestras de machos y hembras adultas y larvas del tercer estadio del desarrollo correspondientes a machos. Se adoptó este diseño ya que se buscaba una caracterización general del transcriptoma, en el que se encontraran representados la mayor cantidad de transcritos posibles y con especial énfasis en las diferencias que pudieran atribuirse al sexo. Esto último en referencia especial a la búsqueda de genes candidatos a estar involucrados en la cascada de determinación del sexo. Dado que no es posible distinguir el sexo de los estadios inmaduros, solamente es viable obtener machos que provendrán de hembras vírgenes (solo producen huevos sin fecundar correspondientes a machos). Treinta individuos adultos de cada sexo y seis larvas del tercer estadio del desarrollo de *D. longicaudata* fueron las condiciones elegidas finalmente ya que se obtuvo una alta calidad e integridad de ARN siguiendo los protocolos y procedimientos detallados en la sección 1. El protocolo de extracción de ARN utilizado en esta sección fue muy similar al ya descrito en la sección 1.1.1. La finalidad en este caso fue la obtención de ARN total que se utilizó para la construcción de bibliotecas de ADNc que fueron posteriormente secuenciadas. Los resultados de integridad evaluada en gel desnaturalizante y de calidad obtenidos espectrofotométricamente (NanoDrop ND1000) obtenidos para las muestras enviadas a secuenciar se muestran en la Figura 14.



**Figura 14.** Corrida electroforética en gel de agarosa 1% (p/v) de muestras de ARN obtenidas a partir de individuos adultos y larvas del tercer estadio del desarrollo de *D. longicaudata*. En la calle 1 se sembró marcador molecular (1000pb y 750pb); ARN total extraído de: larvas del tercer estadio de machos calles 2 y 3; calle 4: hembras adultas y calle 5: machos adultos.

**Tabla 4.** Calidad de extracción de ARN para la obtención de bibliotecas de ADNc. Medidas de calidad de ARN obtenidas espectrofotométricamente (Nanodrop2000).

Muestra	ug/ul	260/280	260/230
1.L3 (10 por muestra)	5,80	1,88	2,10
2.L3 (10 por muestra)	5,90	1,80	2,00
Hembras (30 por muestra)	5,90	2,31	2,20
Machos (30 por muestra)	5,70	2,16	2,05

### 2.1.2. Pirosecuenciación 454 y ensamblado del transcriptoma de *D. longicaudata*

Mediante la técnica de pirosecuenciación 454 (Roche), se generaron un total de 175,7 Mpb de secuencias correspondientes al transcriptoma total de *D. longicaudata*. Esta cantidad de pares de bases se distribuyó entre las 469766 lecturas crudas que se obtuvieron. Como se muestra en la Tabla 8, presentaron una longitud promedio de 376pb. Los datos de secuenciación y

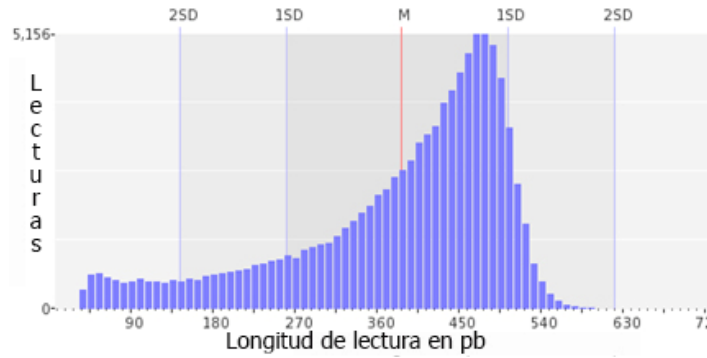
ensamblado se encuentran en el repositorio de *D. longicaudata* (<http://bioinformatica.inta.gov.ar/ATGClongicaudata/>) del INTA Castelar (próximo a estar disponibles en la base de datos Transcriptome Shotgun Assembly, TSA, del NCBI).

Luego de aplicar filtros de secuencias correspondientes a adaptadores, cebadores y secuencias de baja calidad, fue necesario remover 956 lecturas, de esa manera se obtuvieron 468810 secuencias de alta calidad. La cantidad total de lecturas de calidad correspondió al 99,8% del total de las lecturas originales, representando aproximadamente 174,3 Mpb del total para las tres bibliotecas analizadas (Hembras adultas, machos adultos, larvas estadio 3 de macho).

Previo a la secuenciación se construyeron las tres bibliotecas correspondientes a las muestras previamente descritas. Para cada una de ellas se obtuvieron parámetros generales que permitieron conocer la calidad de las secuencias que conforman las mismas. Comenzando con el caso de la biblioteca de hembras, se obtuvieron 98498 lecturas totales que correspondieron a 37,4Mpb. La probabilidad de encontrar una base en la secuencia que no haya podido ser identificada (N) fue de al menos una N cada 11822 secuencias y el máximo de bases sin identificación para una misma secuencia fue del 23%. Por otro lado, la presencia de bases pertenecientes a las secuencias usadas en las etiquetas utilizadas para secuenciar fue menor al 1%. Asimismo, la distribución de calidad de las bases fue buena (Figura 15 y Tabla 5).

**Tabla 5.** Parámetros generales de la biblioteca de hembras.

Distribución de longitudes		Contenido GC	
Longitud promedio	378,01 ± 119.92 pb	Promedio	43,24 ± 8,28 %
Longitud mínima	23 pb	Mínimo	0 %
Longitud máxima	882 pb	Máximo	76 %
Rango de longitudes	860 pb	Rango	77 %
Longitud modal	478 pb con 601 secuencias	Modo	47 con 6794 secuencias



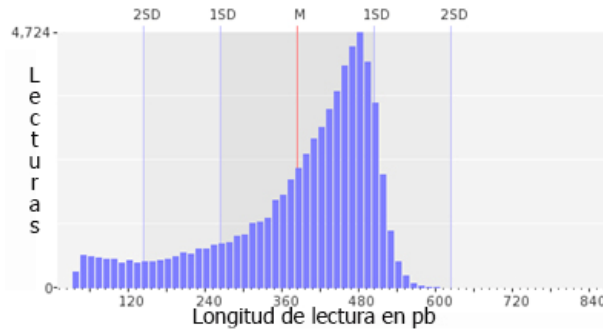
**Figura 15.** Distribución de longitudes de lecturas de la biblioteca de hembras. Eje y: lecturas por bin; eje x: longitud de lecturas en pares de bases (9 por bin).

En el caso de la biblioteca de machos las lecturas totales obtenidas fueron 65892, abarcando un total de 24,9Mpb. Del total de secuencias obtenidas, 7490 secuencias presentaron al menos 1 nucleótido sin identificar (N). El máximo de bases sin identificación para una misma secuencia fue del 10%. La presencia de bases pertenecientes a adaptadores utilizados en el proceso de secuenciación dentro de las secuencias también en este caso fue menor al 1%. Como puede observarse en los gráficos de distribución de calidad de bases no se observan sesgos (Figura 16 y Tabla 6).

**Tabla 6.** Parámetros generales para la biblioteca de machos.

Distribución de longitudes		Contenido GC	
Longitud promedio	377,55 ± 122,49pb	Promedio	42,71 ± 8,43 %
Longitud mínima	18pb	Mínimo	0 %
Longitud máxima	1181pb	Máximo	76 %
Rango de longitudes	1164pb	Rango	77 %
Longitud modal	479pb con 417 secuencias	Modo	47 % with 6794 sequences



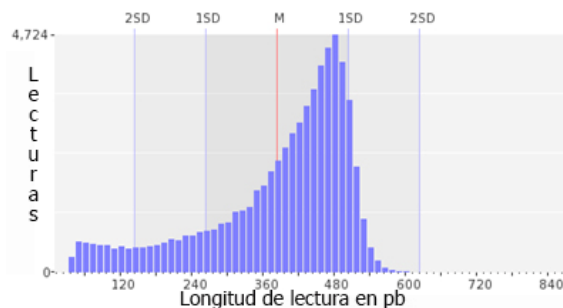


**Figura 16.** Distribución de longitudes de lecturas de la biblioteca de machos. Eje y: lecturas por bin; eje x: longitud de lecturas en pares de bases (12 por bin).

En el caso de la biblioteca de larvas 3 las lecturas totales obtenidas fueron 304926, siendo la que mayor cantidad de bases contuvo (113,5 Mpb). En este caso, la probabilidad de ocurrencia de bases sin identificación fue de al menos una cada 36089 secuencias. El máximo de bases sin identificación para una secuencia fue del 21%. La presencia de bases pertenecientes a las etiquetas utilizadas para secuenciar fue menor al 1% y la distribución de calidad de las bases fue buena (Figura 17 y Tabla 7).

**Tabla 7.** Parámetros generales para la biblioteca de larvas 3 de machos.

Distribución de longitudes		Contenido GC	
Longitud promedio	372,06 ± 128,42 bp	Promedio	43,53 ± 9,78 %
Longitud mínima	21pb	Mínimo	0 %
Longitud máxima	904pb	Máximo	77 %
Rango de longitudes	884pb	Rango	78 %
Longitud modal	480pb con 1088 secuencias	Modo	47 % with 20129 sequences



**Figura 17.** Distribución de longitudes de lecturas de la biblioteca de larvas 3 de machos. Eje y: lecturas por bin; eje x: longitud de lecturas en pares de bases (10 por bin).

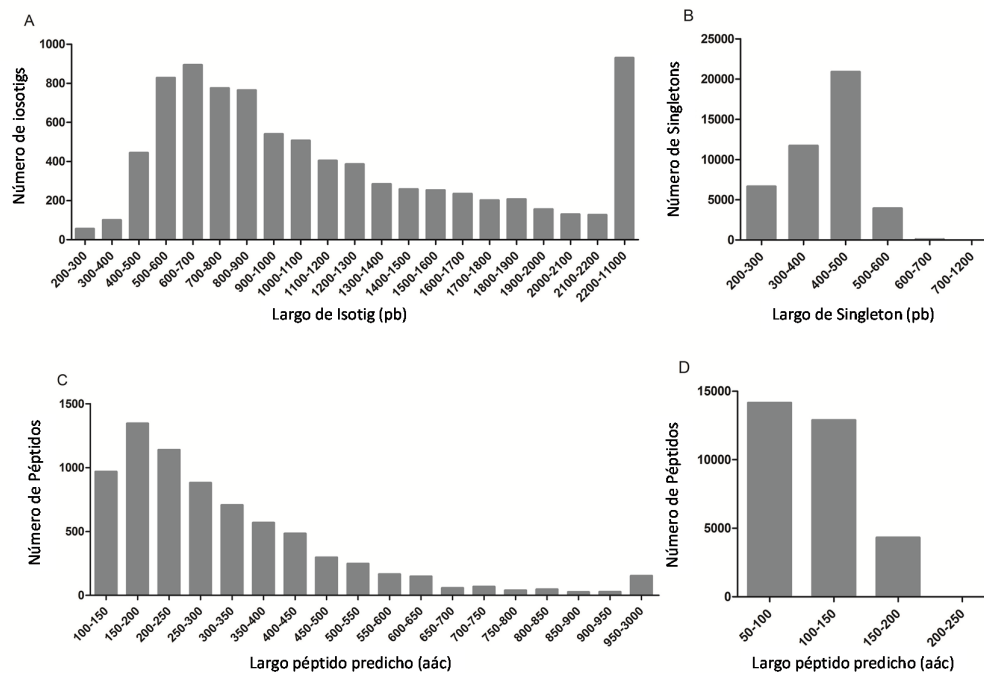
La totalidad de las secuencias de alta calidad obtenidas a partir de las tres bibliotecas de ADNc se agruparon para llevar a cabo el ensamblado *de novo* del transcriptoma de la especie utilizando el software Newbler v2.6 assembler (Roche, IN, USA). Estas secuencias, se ensamblaron en 8483 isotigs (que son equivalentes a transcritos únicos de ARNm) de más de 200pb. Adicionalmente, los isotigs que se originaron como variantes de una misma secuencia, se agruparon en 7295 isogrupos (los isogrupos son equivalentes a un locus genómico). Estas variantes de isotigs que se agruparon en el mismo isogrupo reflejan potenciales variantes de splicing alternativo del ARNm (Tabla 8). Es de particular interés detectar estas variantes para el estudio de las vías regulatorias involucradas en la determinación del sexo ya que, como se ha descrito anteriormente, los genes actualmente caracterizados en organismos modelo son regulados por esta vía.

Alrededor de 47557 lecturas que no pudieron ser ensambladas en ninguno de los isotigs, se ensamblaron utilizando el algoritmo CD-HIT- 454 para eliminar secuencias redundantes dejando 43203 secuencias únicas de una longitud mayor a 200pb. Estas secuencias se denominan singletons y representan al igual que los isotigs secuencias correspondientes a transcritos únicos de ARNm.

En total se logró ensamblar 51686 secuencias no redundantes o unigenes (singletons más isotigs representando secuencias únicas de ARNm) (Tabla 8). La longitud de los isotigs fue desde 200 a 10956pb, con una longitud promedio de 1271,5pb (Figura 18A). Más del 67,2% de los isotigs presentaron una longitud entre 560 y 1300pb y el 93,5% de las bases ensambladas se incorporaron en isotigs de más de 600pb de longitud. La longitud promedio de los isotigs ensamblados para *D. longicaudata* (1271,5pb) fue mayor a la lograda en otras especies de

artrópodos (Vera et al., 2008; Zagrobelny et al., 2009; Pauchet et al., 2010;). Analizando de igual forma las secuencias ensambladas como singletons, se obtuvo que la distribución de longitudes de los 43203 singletons estuvo en el rango de 200 a 1181pb, con un promedio general de longitud de 401,4pb (Figura 18B). La longitud del 99,8% de los singletons fue menor a 600pb.

Asimismo se analizó la distribución de longitudes de los péptidos predichos a partir de las secuencias ensambladas como isotigs y singletons. El total de péptidos que se lograron predecir a partir de las secuencias de nucleótidos fue de 38685, 7348 a partir de isotigs y 3133 a partir de singletons (Tabla 8). Las longitudes promedio fueron 324,5 y 109,9 aminoácidos, respectivamente (Figura 18C y 18D).



**Figura 18.** Distribución de longitudes de los Isotigs (A); Singletons (B); Péptidos predichos a partir de los Isotigs (C) y péptidos predichos a partir de los Singletons (D).

## 2.2. Obtención de marcadores moleculares a partir de ARN

### 2.2.1. Diversidad Genética: predicción de marcadores

Como una de aplicaciones de las técnicas de secuenciación masiva NGS, se obtuvieron para la especie una batería de marcadores moleculares *in silico*. La generación de esta vasta cantidad de información es de potencial utilidad en la genética de poblaciones y en el análisis de calidad de producción de insectos.

#### 2.2.1.1. Obtención *in silico* de Repeticiones de Secuencia Única (SSR)

Haciendo uso del software MISA, se identificaron y caracterizaron los motivos de SSR en el transcriptoma de *D. longicaudata* para ser usados como potenciales marcadores moleculares específicos para este parasitoide. Se identificaron 1233 SSR putativos en los 8515 isotigs y 3441 SSR putativos en los 43203 singletons. Las frecuencias para la ocurrencia de los SSR considerando múltiples SSR en un transcripto, fue del 15% para los isotigs y de 8% para los singletons. En el caso de los isotigs, el porcentaje de ocurrencia de SSR en *D. longicaudata* fue más elevado que los previamente reportados para *Spalangia endius* Walker (Hymenoptera: Pteromalidae) (8.51%) (Zhang et al., 2014) y para *Laodelphax striatellus* (Fallén) (Homoptera: Delphacidae) (7%) (Zhang et al., 2010). Un total de 982 (12%) de los isotigs presentaron al menos un SSR. Es de interés avanzar en el diseño de cebadores específicos a fin de amplificar las regiones que contienen a estos SSR para evaluar su utilidad real como marcadores moleculares. Por este motivo se analizó este aspecto para los SSR obtenidos. Para 693 SSR, representando el 70,6% del total, se halló que presentaban secuencias flanqueantes lo suficientemente largas como para diseñar cebadores únicos específicos para amplificar el marcador apropiadamente. Lo mismo se analizó para el caso de los singletons y un total de 3441 presentaron al menos un SSR, de entre los cuales 2082 SSR (60,5%) permitieron el diseño de cebadores únicos específicos (disponible en <https://mega.nz/#!PJ1TCBiS>).

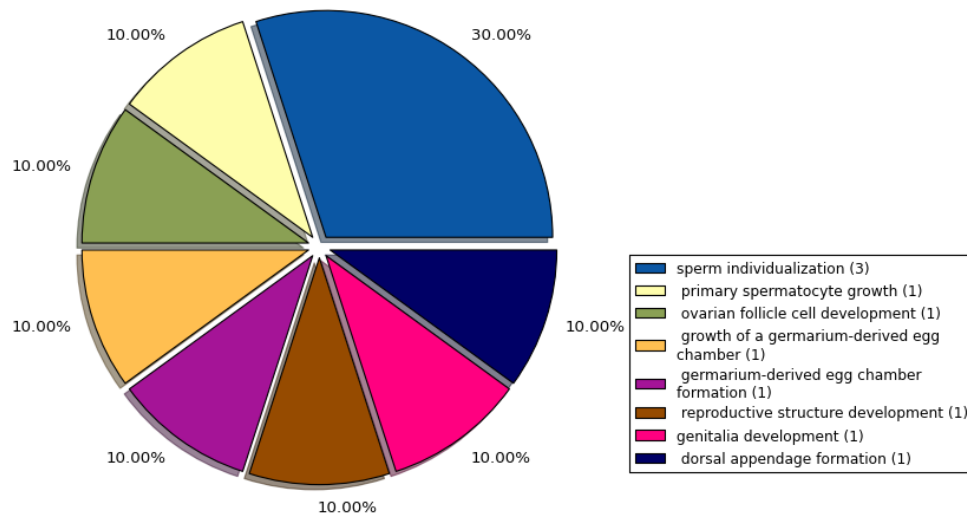
#### 2.2.1.2. Caracterización de motivos y distribución de microsatélites

El tipo más frecuente de microsatélite se correspondió con los motivos tetramérico (38,8%), dimérico (24,3%) y trimérico (14,9%), apareciendo las repeticiones penta, hexanucleotídica y repeticiones compuestas, en mucha menor frecuencia (6,6%; 4,2% y 11% respectivamente, <https://mega.nz/#!PJ1TCBiS>). Estos resultados contrastan con los encontrados en *S. endius* y *L. striatellus*, donde los di y trinucleótidos fueron las repeticiones más frecuentes (Zhang et al., 2010, 2014).

Las combinaciones de motivos de SSR pueden ser agrupados en clases únicas basándose en la complementariedad de bases del ADN. Los tetranucleótidos fueron agrupados en 27 clases y los números de clases únicas para las repeticiones de di y trinucleótidos fueron 4 y 10, respectivamente (Anexo x). La combinación más frecuente para los tetranucleótidos fueron tres repeticiones de AAAT/ATTT; cinco repeticiones de AG/CT para dinucleótidos y cinco repeticiones de AAT/ATT para los trinucleótidos (<https://mega.nz/#!PJ1TCBiS>). Estos motivos son principalmente repeticiones AT, esto concuerda con el hecho de que los genomas de Hymenoptera se caracterizan por ser especialmente ricos en A-T (Stolle et al., 2013).

La información acerca de la identidad (ID), motivo de la repetición, longitud de la repetición, secuencias de cebadores, posiciones de los cebadores *forward* y *reverse* y longitud de fragmento de amplificación esperados se encuentran disponibles en <https://mega.nz/#!PJ1TCBiS>.

La distribución de SSR se analizó también específicamente entre los transcriptos vinculados a determinación del sexo como subgrupo dentro de aquellos anotados como procesos biológicos (Tabla 9). Se encontraron diez marcadores potenciales asociados a estos transcriptos (Figura 19). Todos los marcadores identificados tienen potencial para ser usados en mapeo genético, análisis parental, flujo génico y genética de poblaciones.



**Figura 19.** Distribución de los marcadores en las secuencias identificadas como involucradas en determinación del sexo y funciones reproductivas asociadas por medio de GO (Tabla 9).

## 2.3. Anotación funcional: identificación y análisis de secuencias

Con la finalidad de asignar una función putativa y una ontología génica (GO) a cada unigen, se sometió el conjunto de secuencias a una búsqueda de similitud por medio de BLASTx contra la base NR de proteínas del National Center for Biotechnology Information (NCBI) seguida de una búsqueda por medio de BLAST2Go e InterProScan para completar la anotación (Gish and States, 1993; Zdobnov and Apweiler, 2001). Una vez obtenidos estos resultados, se combinaron según su anotación por medio de GO.

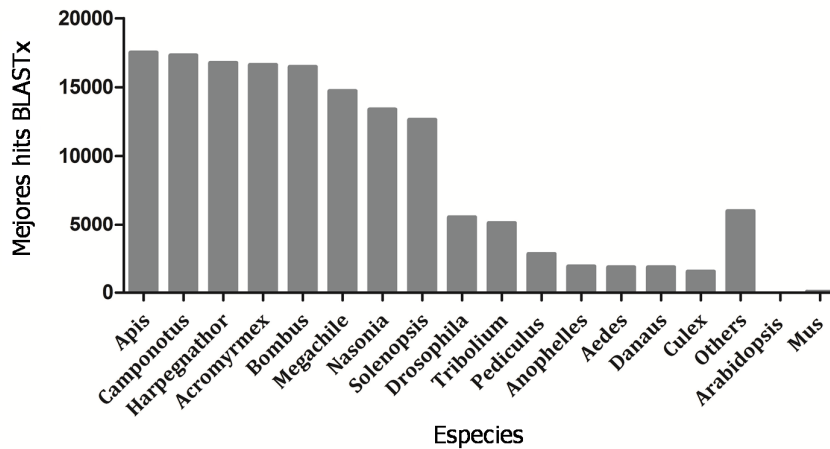
Del total de secuencias analizadas, isotigs y singletons, 7021 isotigs (82,7%) y 20227 singletons (46,8%) presentaron homología suficiente para ser identificados por medio de BLASTx. (Tabla 8). La frecuencia de isotigs/singletons anotados por este medio fue de 48,6%, presentando un mayor porcentaje que los previamente reportados para ensamblados *de novo* de transcriptomas de especies relacionadas que no cuentan con datos de secuenciación para sus genomas. En estos casos los valores de anotación reportados son de entre el 20 y el 40% (Badouin et al., 2013; Nishimura et al., 2012).

**Tabla 8.** Resumen de parámetros de calidad del transcriptoma de *D. longicaudata*

Número total de lecturas de alta calidad	468810
Lecturas de alta calidad de la muestra de larvas	304174
Lecturas de alta calidad de la muestra de hembras	98801
Lecturas de alta calidad de la muestra de machos	65835
Longitud promedio (bp)	376
Número total de contig/isotigs	8483
Longitud promedio de contig/isotigs (bp)	1271.5
Número total de singletons	43203
Longitud promedio de singletons (bp)	401.4
Número total de peptides predichos	38685
Longitud promedio de péptido (aminoácidos)	217.2
Total de isotigs anotados (BLASTx)	7021
Total de singletons anotados (BLASTx)	20227

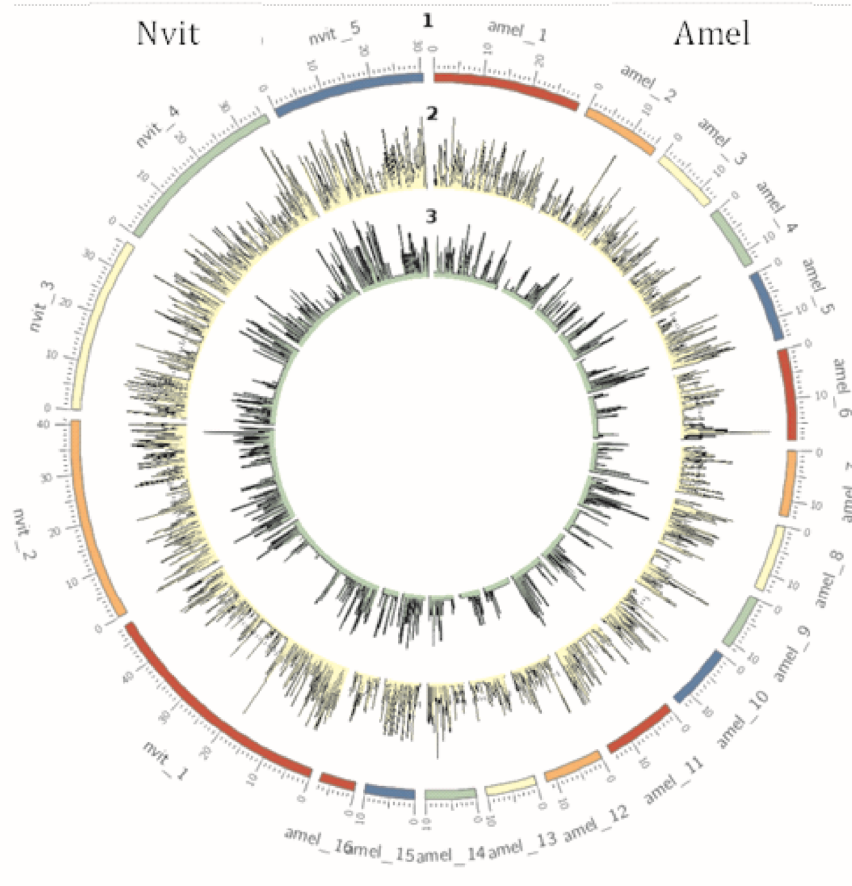
A fin de obtener una visión general de las identidades de las secuencias representadas en el transcriptoma de *D. longicaudata*, se realizó una búsqueda sobre la base de datos del parasitoide de aquellas especies que mostraban mayor homología por medio de BLASTx. La

mayoría de las secuencias de *D. longicaudata* analizadas presentaron, como era de esperarse, homología con otros himenópteros. Se observó principalmente una alta similitud con secuencias de *Apis* (11,5%), *Camponotus* (11,4%) y *Haemaphysalis* (11%). La distribución de mejores coincidencias de BLASTx con especies relacionadas se muestra en la Figura 20.



**Figura 20.** Distribución de los hits de BLASTx de la base de datos de *D. longicaudata* generada a partir de las tres bibliotecas de ADNc. Se incluyen *Mus musculus* y *Arabidopsis thaliana* como representativos de especies no relacionadas.

A modo de evaluación de la calidad de los datos de transcriptómica obtenidos para *D. longicaudata*, se llevó a cabo una comparación general de distribución de transcritos con los genomas de dos especies modelo relacionadas dentro del orden Hymenoptera: *A. mellifera* (Amel) y *N. vitripennis* (Nvit) (Figura 21). Al observar la comparación, se vio una distribución uniforme de los transcritos del parasitoide a lo largo de ambos genomas, encontrando representantes en casi todos los cromosomas de las especies modelos.

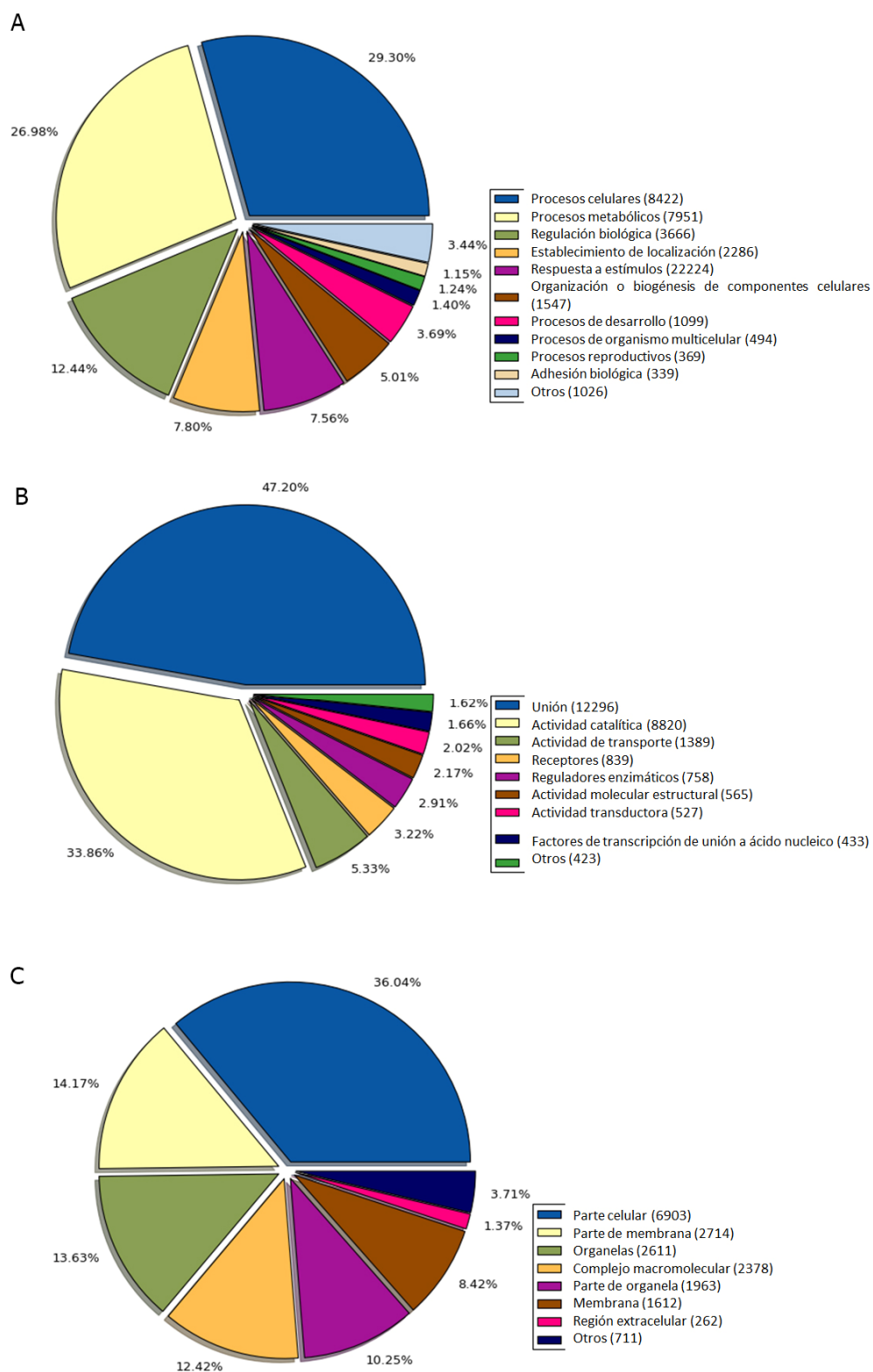


**Figura 21.** Comparación del transcriptoma de *D. longicaudata* con los genomas y transcriptomas de himenópteros relacionados. *N. vitripennis* (Nvit) y *A. mellifera* (Amel). En el primer anillo se observan los dos genomas de las especies modelo. En el segundo anillo se encuentran enfrentados los transcriptomas de Nvit y Amel, siendo el tercer anillo la comparación de *D. longicaudata* con ambos transcriptomas.

Las secuencias se anotaron haciendo uso de los términos de Ontología Génica (GO) y las categorías de la Comisión de Enzimas (i.e. números EC). Se asignaron términos GO a un total de 16930 transcriptos, 4483 isotigs (52,8%) y 12447 singletons (28,8%). El análisis por medio de InterProScan resultó en 6535 coincidencias significativas para proteínas traducidas desde los isotigs (77,0%) y 20979 para sigletons (48,6%). Los términos GO asignados se dividen en tres categorías principales: procesos biológicos, funciones moleculares y componentes celulares. Los procesos biológicos abarcan todos los procesos involucrados en el funcionamiento de los niveles de organización: células, tejidos, órganos y organismos. Se entiende en este caso proceso como colección de eventos moleculares que tienen un principio y final definidos. Las funciones moleculares se entienden como actividades elementales tales como la catálisis o aquellas



proteínas de unión, describen acciones llevadas a cabo por productos génicos a nivel molecular. Al producto de un gen determinado pueden asignársele varias funciones moleculares. Tal vez el caso de los componentes celulares es el más claro, refieren a las partes de la célula o el ambiente extracelular en donde se encuentra localizado un producto génico. Estos productos génicos pueden estar localizados en una o más partes de la célula y su ubicación podrá ser altamente específica, por ejemplo al formar parte de un complejo macromolecular en particular, que es una asociación estable y persistente de macromoléculas que funcionan en conjunto. A partir de los términos de GO asignados a isotigs y singletons, 13142 términos se asignaron a “Procesos Biológicos”, 17188 términos se asignaron a “Funciones moleculares” en referencia a procesos elementales del metabolismo y 9911 términos a “Componentes celulares” (Figura 22). Continuando con la asignación funcional, a través del software BLAST2GO, se pudo evidenciar que de entre los 11 procesos biológicos diferentes, la mayoría de los transcriptos fueron asignados a procesos celulares (8422) o a procesos metabólicos (7951) (22 A). Las subcategorías representadas mayoritariamente entre las funciones moleculares fueron de unión (12296), actividad catalítica (8818) y transportadoras (1389) (22 B). En la categoría de componentes celulares, todos los transcriptos allí anotados fueron clasificados en ocho subcategorías, siendo las más representadas: partes celulares (6903), partes de la membrana celular (2714) y componente de complejos macromoleculares (1136) (22 C).



**Figura 22.** Anotación funcional de los transcritos ensamblados. A. Procesos biológicos, 13142 términos; B. Funciones moleculares (procesos regulatorios elementales) 17188 términos y C. Componentes celulares, 9911 términos.

Para una caracterización más profunda del transcriptoma de *D. longicaudata*, se realizó una búsqueda de aquellas secuencias que estuvieran asociadas específicamente a las funciones involucradas en la determinación del sexo y a los procesos del desarrollo implicados en la reproducción y la determinación sexual. Estos términos se encuentran clasificados dentro de la categoría de procesos biológicos (Tabla 9). Un total de cuatro secuencias fueron anotadas como relacionadas a la determinación del sexo (GO:0007530). En un nivel más específico, estas secuencias fueron anotadas bajo funciones primarias de determinación del sexo (GO:0007538), línea somática (GO:0007539) y línea germinal (GO:0007542; específicamente línea germinal femenina GO:0018992; GO:0019099 y GO:0030237). Se hallaron asimismo diez secuencias relacionadas a procesos reproductivos entre las cuales cabe destacar a los isotigs00949 y 00948 que representan variantes del mismo transcripto. Esto podría reflejar potenciales variantes de “splicing” alternativo del ARNm asociadas al sexo (Tabla 9).

**Tabla 9.** Secuencias involucradas en la determinación del sexo

Nombre	GO	Secuencia	Marcadores y posición
Determinación del sexo	GO:0007530	HWRCMJ02HFI99; HWRCMJ02HFCXO; HWRCMJ02FPEMS; HWRCMJ02IL9LL	
Determinación del sexo primaria	GO:0007538	HWRCMJ02IL9LL; HWRCMJ02HFCXO; HWRCMJ02FPEMS	
Determinación del sexo primaria, soma	GO:0007539	HWRCMJ02HFI99	SSR_801; 326-355 SSR_802: 160-171
Determinación del sexo primaria; línea germinal	GO:0007542	HWRCMJ02IL9LL	
Diferenciación sexual	GO:0007548	HWR9CMJ02FUVP3; HWR9CMJ02FK4N5; HWR9CMJ02H4520	
Formación del huevo derivada del germario	GO:0008103	Isotig02368	SSR_537; 936-951
Individualización del esperma	GO:0031011:	isotig00948	SSR_280; 1083-1102
	GO:0031011	isotig00949	SSR_279; 1083-1102 SSR_333; 2120
	GO:0008407	isotig04171	SSR_777; 592-601
Formación del apéndice dorsal	GO:0048814:	isotig02537	SSR_561; 1242-1251

Diferenciación de células del folículo ovárico	GO:0045879	isotig03388	SSR_691; 52-69 SSR_692; 1436-1450
Crecimiento primario del espermatocito	GO:0030307	isotig03510	SSR_708; 197-206

Determinación del sexo en *Diachasmimorpha*  
*longicaudata*

### 3. Análisis de genes candidatos

#### 3.1. Validación de resultados de secuenciación

La obtención masiva de datos por pirosecuenciación 454 generó una enorme cantidad de información que debe ser validada a fin de determinar un criterio válido y confiable para la interpretación de datos. A fin de confirmar las tendencias de los perfiles de expresión caracterizados para los sexos, se seleccionó y analizó por RT-qPCR un grupo de transcritos.

##### 3.1.1. Puesta a punto y determinación del rango de linealidad de la RT-qPCR: Genes de Referencia y Genes Candidatos.

###### 3.1.1.1. Puesta a punto de la RT-qPCR y construcción de las curvas de calibración

La PCR en tiempo real (RT-qPCR) es una técnica robusta que permite cuantificar ácidos nucleicos con gran precisión gracias a su alta sensibilidad y al cuidadoso análisis estadístico que requiere. Si bien existe una gran variedad de aplicaciones y métodos de análisis de resultados en relación a la aplicación que se requiera, aun no se cuenta con un criterio formal de validación de la expresión diferencial. Por este motivo, es necesario tomar criterios estrictos a fin de obtener resultados confiables y reproducibles. Para este sistema fue necesaria la elección y puesta a punto de Genes de Referencia (GR).

###### 3.1.1.1.1. Elección de GR

Se seleccionaron como GR a la  $\beta$  actina y el *factor de elongación 1  $\alpha$* , en base a datos bibliográficos (Infante et al., 2008) y ensayos previos en *A. mellifera* dentro del grupo de trabajo (Scannapieco et al., 2011). Los cebadores específicos y las identidades de las secuencias se encuentran listadas en la Tabla 11. En ambos casos fue necesario corroborar su aplicabilidad como GR (Hellemans et al., 2007) ya que se trata de un sistema del que no se cuenta con datos previos. Como se observa en el apartado siguiente y en la tabla 11, tanto el rango dinámico como los valores obtenidos para las muestras con ambos GR permitieron su uso.

**Tabla 11.** Parámetros de validez de GR. Coeficiente de variación (CV, aceptable menor a 25%) y M (aceptable menor a 0,5).

Gen	CV	M
<i><math>\beta</math>-actina</i>	13,7	0,13
<i>Factor de elongación 1 <math>\alpha</math></i>	12,4	0,12

#### 3.1.1.1.2. Elección de los GI

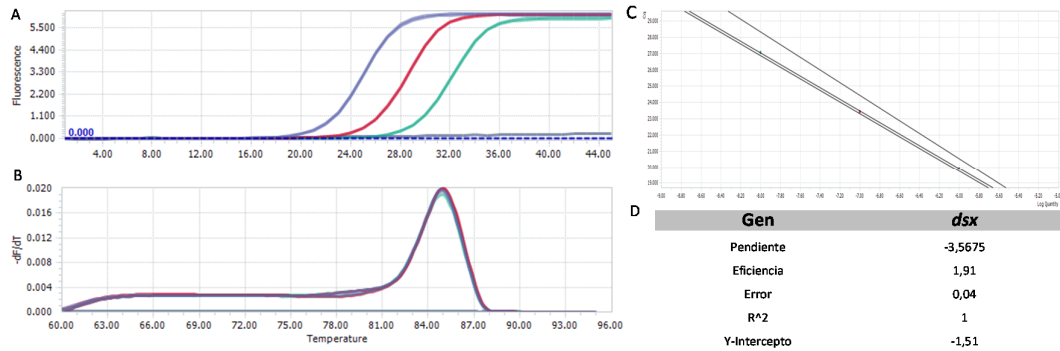
En esta sección se busca evaluar la validez de los resultados obtenidos por medio de transcriptómica en relación a la expresión sexo-específica. Los genes de interés (GI) seleccionados para evaluar las tendencias de expresión fueron tres de los isotigs que mostraron diferencias significativas entre los sexos por medio de pirosecuenciación según los criterios adoptados previamente descritos. Los isotigs seleccionados se muestran en la Tabla 10, también pueden observarse los valores de eficiencia de amplificación de los cebadores diseñados sobre las secuencias de estos transcritos.

#### 3.1.1.2. Curvas de calibración

A fin de determinar el rango dinámico de las diferentes reacciones y la eficiencia de los pares de cebadores utilizados se construyeron curvas de calibración para cada uno de ellos. El material usado como molde para estas curvas se obtuvo a partir de diluciones seriadas del producto purificado de PCR correspondiente a cada par de cebadores. En los diferentes casos fueron utilizadas al menos cuatro diluciones entre el rango de -2 y -8 para construir la curva y determinar el rango dinámico de la reacción donde es válida la cuantificación. Una vez construidas las curvas de calibración se analizaron los picos de desnaturalización de los productos de PCR y se procedió a la secuenciación de los mismos para confirmar la identidad de los productos.

En todos los casos se obtuvo un rango dinámico que incluía las muestras permitiendo su cuantificación con un mínimo margen de error, así como una buena eficiencia de reacción, entre 90 y 110% que son los márgenes aceptables (Kubista et al., 2006) (Tabla 11).

Del mismo modo que se procedió a evaluar los parámetros de la reacción para los GR, se evaluaron también para todos los Genes de Interés (GI). Los resultados de las eficiencias de amplificación se muestran en la Tabla 10. A modo ilustrativo se muestra en la Figura 23 una de las Curvas de Calibración construidas para el GI *dsx*.



**Figura 23.** Curvas de amplificación de RT-qPCR para el gen *dsx*. Se muestran las curvas de fluorescencia en el tiempo (A); las curvas de desnaturalización (B) y las curvas de calibración (C). Finalmente se muestran en la Tabla D los valores obtenidos para la Curva de calibración mostrada en C.

**Tabla 11.** Cebadores específicos y características de la RT-qPCR.

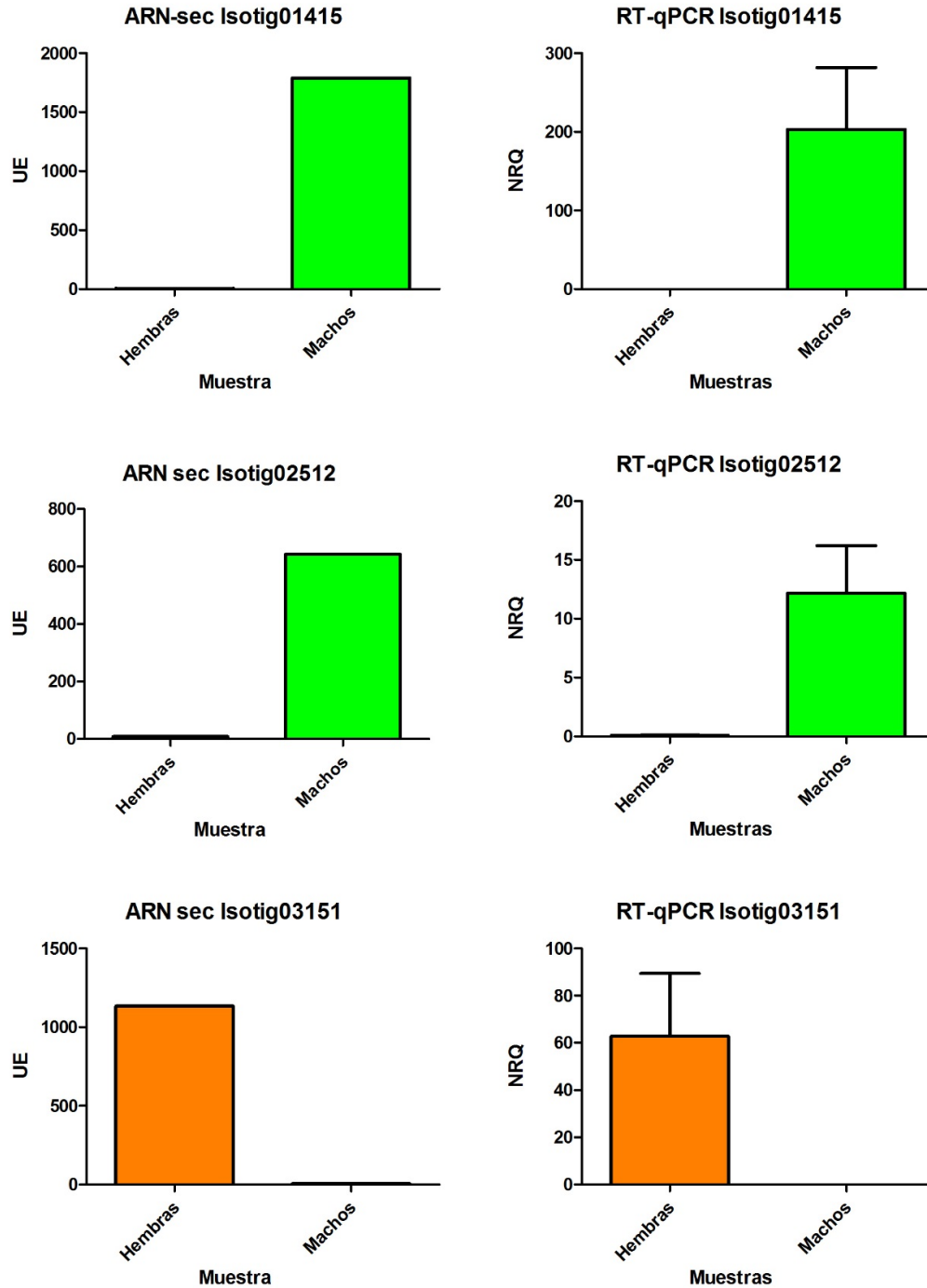
ID	Nombre	Mejor hit en BLAST	Secuencia 5'-3'	Producto (bp)	Eficiencia de amplificación (%)
Isotig009459	<i>β-actina</i>	[NP_511052; XP_001986647]	Fw agcaccagctcctcttgac Rv aacaccatcaccgagtc	182	93.0
Isotig00208	<i>Factor de elongación 1 α</i>	[BAM18878; BAG30769]	Fw tgccttcgttcccatctccg Rv tcaggcatttaccctcagcc	122	90.0
isotig01415	<i>Receptor odorante</i>	[XP_003708550; XP_003494618]	Fw taaggctcgcaaaggcgaat Rv tgcgcgtgtgacgatttca	182	93.0
isotig02512	<i>Esfingomielina fosfodiesterasa</i>	[XM_001122062; XM_003493337]	Fw ccgaggaagggttccgaat Rv cggtcaggtcggttggttg	153	99.0
isotig03151	<i>Myrosinasa 1-like</i>	[XP_001601101; XP_001600108]	Fw aggtgcctggaatgtcagt Rv acggtaatggcgaaacccaa	205	100.3
isotig07202	<i>Feminización 1-like</i>	[NP_001153369; EFN87550]	Fw ttgttcaggcagctaagca Rv gtcacgagtattcaggggca	156	103
isotig06880	<i>Fem 1-like</i>	[EFN89777; EFN60633]	Fw tggttggggaggaaggaaga Rv taaggtcacggtcgcagttg	239	91.3
isotig08283	<i>doublesex</i>	[XP_003700217; ABW99103]	Fw gtcagtaccacagccagctc Rv atgttgtgggactgcggtag	152	89.7



### 3.1.2. Validación de los patrones de expresión por medio de análisis comparativo

Si bien las técnicas de secuenciación masiva NGS generan una enorme cantidad de información y permiten un análisis de estas secuencias, estos resultados deben ser validados por medio de técnicas más precisas y sensibles. En el caso de la pirosecuenciación 454 (Roche) el análisis de expresión diferencial debe ser corroborado para establecer la validez del ensayo seleccionando genes y confirmando las tendencias de expresión. La técnica elegida para llevar a cabo esta validación, por las razones expuestas previamente, es la RT-qPCR.

Los resultados obtenidos por comparación de las dos técnicas se muestran en la Figura 24. Para el isotig02512 se encontró, tanto por pirosecuenciación como por RT-qPCR, que se encuentra sobreexpresado en machos. En el caso de la pirosecuenciación se vio un incremento de 34 veces y alrededor de  $10^5$  veces por medio de RT-qPCR ( $p=5 \times 10^{-07}$ ) (Figura 24). Para el isotig01415, se encontró un incremento de la expresión de 329 veces por medio de la secuenciación de ARN en los machos. Por medio de RT-qPCR se observó el mismo patrón, confirmando la sobreexpresión en machos con un nivel de sobreexpresión de 940 veces ( $p=1.5 \times 10^{-6}$ ) (Figura 24). Por el contrario, para el isotig03151, se observó la tendencia inversa, sobreexpresándose en hembras por 441 veces por medio de secuenciación de ARN y con un nivel de sobreexpresión de 15 veces al evaluarla por la técnica de RT-qPCR ( $p=10^{-6}$ ) (Figura 24). Estos resultados sugieren que al analizar los datos de lecturas de pirosecuenciación de ARN normalizadas, a partir de un determinado umbral de corte, los datos podrán ser usados de manera confiable para la determinación de perfiles de expresión.



**Figura 24.** Comparación entre los resultados de expresión obtenidos por medio de NGS y RT-qPCR. En la columna de la izquierda se observan los resultados de NGS y en la columna de la derecha se observan los resultados obtenidos por RT-qPCR para los mismos transcritos. UE indica las unidades de expresión y NQR es la cantidad relativa normalizada (ver materiales y métodos para una definición más amplia).

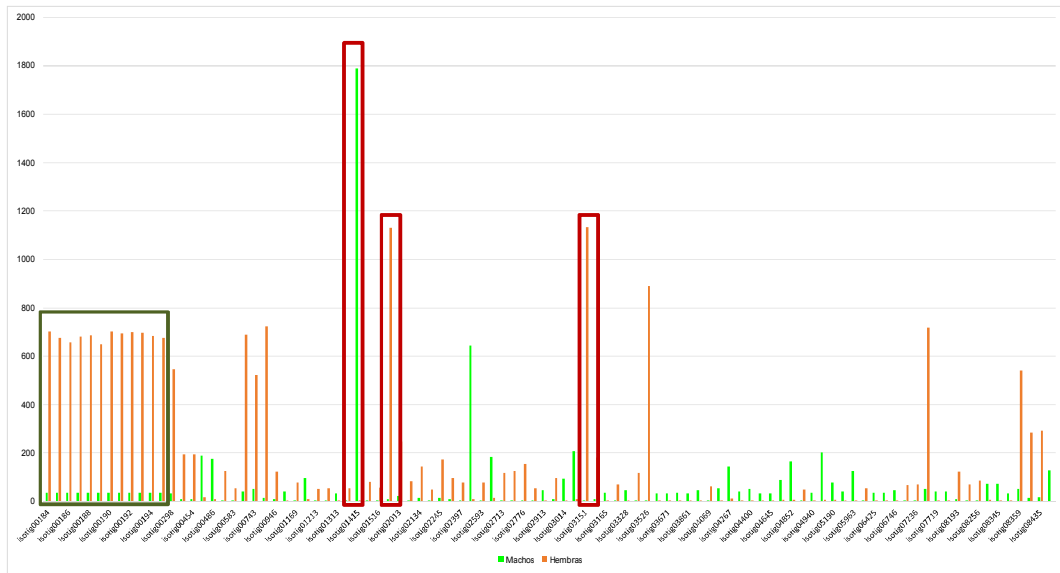
### 3.2. Análisis comparativo del transcriptoma

A partir de las tres bibliotecas de ADNc secuenciadas y ensambladas, se realizó una comparación entre las bibliotecas generadas para hembras y machos adultos de *D. longicaudata* con la finalidad de encontrar las diferencias entre sexos. Esto se llevó a cabo apuntando a descubrir potenciales candidatos a estar involucrados en las vías de determinación del sexo observando las diferencias sexo específicas en los patrones de expresión medidos. Los sexos son morfológicamente distintos y cumplen funciones biológicas diferentes, por lo que en este estudio se evaluaron los genes con los patrones de expresión que presentaron mayor diferencia entre estas dos condiciones. Estas secuencias diferencialmente expresadas se muestran en la Figura 25, evidenciando que existe un mayor nivel de expresión en el caso de las hembras para la mayoría de las secuencias.

Un análisis más detallado de las secuencias que presentaron una diferencia de expresión entre sexos, considerando como diferentes a todas aquellas secuencias que presentaron una diferencia de más de 10 veces, reveló que en total 98 transcriptos mostraban una tendencia a estar diferencialmente expresados entre hembras y machos adultos (Anexo 4). Entre estas secuencias, 59 de ellas se vieron sobreexpresadas en las hembras mientras que 39 secuencias se vieron sobreexpresadas en machos (Anexo 4).

De entre las secuencias sobreexpresadas en hembras puede observarse un grupo de transcriptos que tienen un nivel de expresión similar y pertenecen al grupo de isotigs que abarca desde el isotig00184 hasta el isotig00195; dichos transcriptos conforman el isogrupo00011 que no tiene función asignada. Estos transcriptos se muestran en la Figura 25 dentro del recuadro verde. Para hacer un análisis más profundo de estas secuencias que se agruparon bajo un mismo isogrupo (recordar que es equivalente a un locus génico) se alinearon a fin de evaluar la existencia de diferencias entre ellas, haciendo hincapié en aquellas que pudieran representar variantes de “splicing” de ARNm. Al observar el alineamiento estas secuencias se dividen en dos grupos, mostrando cambios puntuales de bases y una delección de tres bases en la posición 414-416 (Anexo 6). Al hacer una nueva búsqueda por medio de BLASTx, estas secuencias se encuentran en especies relacionadas de himenópteros como *F. arisanus*, *T. pretiosum*, *M. rotundata*, *N. vitripennis* y *A. mellifera*. En todos los casos la secuencia se encuentra registrada como no caracterizada. Sería necesario hacer una evaluación más profunda de este grupo de secuencias en relación a su función biológica dadas las características que presentan y el hecho de que no se encuentran identificadas funcionalmente.

Se pudo estimar de esta manera, que aproximadamente el 1,2% del transcriptoma muestra un cambio en al menos diez veces en su expresión entre sexos, con una mayoría de transcritos sobreexpresándose en hembras. Es interesante observar que pese a las diferencias entre machos y hembras y a las particularidades del ciclo de vida *D. longicaudata*, la mayor parte del transcriptoma se expresa de manera estable entre ambos sexos. Entre estos grupos de transcritos se encuentra una porción de los mismos que no han podido ser anotados, sugiriendo que estos formarían parte del transcriptoma específico de *D. longicaudata* (Anexo 5, Tabla A1).



**Figura 25.** Transcritos expresados diferencialmente entre hembras y machos adultos de *D. longicaudata*. Se muestran en recuadros rojos los transcritos utilizados para corroborar los resultados de transcriptómica por medio de RT-qPCR. En el recuadro verde se muestra un grupo específico de transcritos que se sobreexpresa en hembras y adicionalmente pertenece al isogrupo00011.

### 3.3. Selección de genes de determinación del sexo

De acuerdo con análisis filogenéticos, el sistema ancestral de determinación del sexo dentro del orden Hymenoptera es el de alelos complementarios (CSD) y es una característica reconocida de la reproducción de los himenópteros aunque actualmente no puede determinarse si el estado ancestral fue sl-CSD o ml-CSD (Asplen et al., 2009). El mecanismo CSD comprende la acción de varios genes en cascada como *csd/tra*, *fem*, y *dsx*; y si bien es mayoritario entre los himenópteros, no es el único. Recientemente se ha estudiado también el Sistema de

Determinación del Sexo por medio del Efecto de “Imprinting” Genómico Materno (MEGISD), presente en especies como *N. vitripennis* (Verhulst et al., 2010, 2013). En función de estos sistemas previamente caracterizados y los datos obtenidos para *D. longicaudata* se realizó una selección de transcriptos candidatos a estar involucrados en la determinación del sexo. La selección se llevó a cabo según dos criterios. En primer lugar se hizo una búsqueda de todos aquellos transcriptos que presentaron una diferencia en la cantidad de lecturas entre la biblioteca de ADNc de hembras y la de machos adultos (diferencia de 10 o más veces). En segundo lugar, se seleccionaron aquellos transcriptos que presentaron homología con los genes previamente caracterizados en otras especies de artrópodos como involucrados en los mecanismos de determinación del sexo: *fem* y *dsx*.

### 3.3.1. Caracterización de transcriptos previamente obtenidos: *csd*

Retomando lo discutido en la sección 1 acerca de los genes candidatos a estar involucrados en los mecanismos de determinación del sexo para *D. longicaudata*, se buscó identificar ahora sobre la base de datos generada por transcriptómica para el parasitoide, los fragmentos secuenciados en ADN y ADNc que inicialmente se creía que formaban parte del transcripto correspondiente al gen *csd*. Al comparar estas secuencias con la base de datos para el parasitoide, se encontró que los fragmentos considerados previamente como Región conservada y Región variable formaban parte de dos transcriptos diferentes sin relación alguna con el gen *csd*. En el caso de la región conservada mostró un 100% de homología con el isotig002035, que tuvo como mejor hit de BLASTx una metalo-endopeptidasa 1 de *A. mellifera*; mientras que para la región variable tuvo una homología de 100% con el isotig02265, cuyo mejor hit en BLASTx fue el Miembro 4 de la subfamilia G del cassette de unión a ATP de *Fopius arisanus* (Hymenoptera: Braconidae) [XP\_011299286.1].

### 3.3.2. Caracterización de secuencias de genes candidatos identificados por transcriptómica

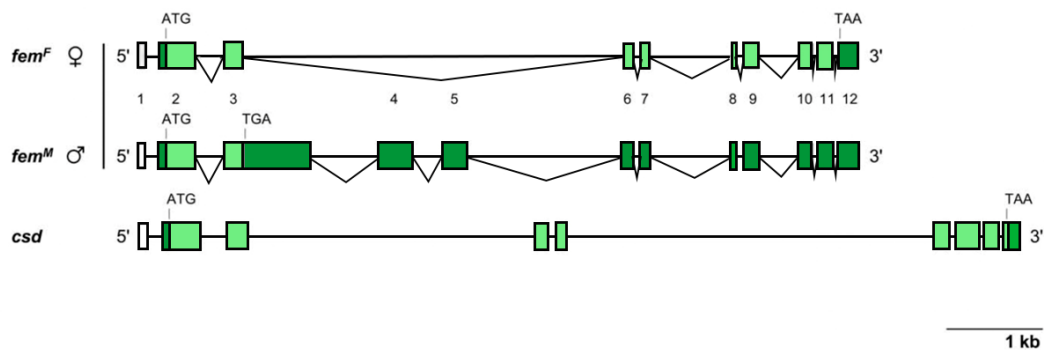
En una segunda aproximación, se encontraron secuencias homólogas a genes que ya se han identificado como involucrados en la determinación del sexo (Beye et al., 2003; Hasselmann et al., 2008; Werren et al., 2009; van de Zande and Verhulst, 2014). A través de la búsqueda por medio de BLASTx, se centró la atención en *fem*, regulador río arriba conservado, [NP\_001153369; EFN87550; NM001159897.1; XM008553448.1; XM003704454.1;

XP008559611.1; EFN89777.1; XP003703392.1; XP006608847.1; EFN60633] y en el efector altamente conservado *dsx* [gi|383849166; gi|383849165; XP003395766; ABW99103.1; NP001128407.1; XP\_003700217]. De este modo se hallaron secuencias de interés en el set de datos ensamblados para *D. longicaudata* que presentaron homología a los genes mencionados a nivel proteína (BLASTx): los isotigs 07202/06220 y 06880/06880b que presentaron homología funcional con *fem*; y los isotigs 08283/00774/00775 que presentaron homología funcional con *dsx*.

Para la correcta identificación de los transcritos mencionados, se continuó con la caracterización de los mismos a fin de obtener la mayor porción de secuencia posible.

### 3.3.2.1. *Feminizer*

El gen *feminizer* (*fem*), ortólogo al gen *transformer* (*tra*), ha sido ampliamente caracterizado en insectos holometábolos como el gen ancestral encargado del control del desarrollo sexual y de la regulación del splicing alternativo del efector *dsx*. Este gen es el progenitor conservado del gen *csd* que adquirió una nueva función en el sistema de determinación del sexo por alelos complementarios presente en más de 60 especies de himenópteros (Hasselmann et al., 2010), entre ellas, *D. longicaudata* (Carabajal Paladino et al., 2015).



**Figura 26.** Esquema del gen *fem* y *csd* en *A. mellifera*. Se muestran los exones en forma de cajas: en verde y los intrones en líneas negras. En el caso del gen *fem* se muestra el resultado del “splicing” alternativo para los sexos (Gempe et al., 2009).

Como resultado de la búsqueda por similitud (BLASTx) al gen *tra/fem* a partir de las secuencias obtenidas por transcriptómica, se identificaron cuatro secuencias que presentaron un alto grado de homología a nivel dominio funcional con las proteínas TRA/FEM caracterizadas en otros himenópteros. Los dominios conservados identificados se obtuvieron a partir de la base de datos de proteínas PDB (<http://www.wwpdb.org/>) y se utilizaron en los alineamientos realizados entre las secuencias de *D. longicaudata* identificadas como variantes de *fem* y las mejores coincidencias obtenidas por medio de BLASTx.

Si bien es posible ubicar pares de isotigs con un porcentaje alto de homología sobre variantes de las secuencias de la proteína FEM, para las dos variantes halladas existe un fragmento común sobre el dominio funcional conservado correspondiente a la familia de las Ankrinas, específicamente Ank\_4. En esta zona podemos ubicar tanto al isotig06880 como al isotig06260 que a su vez presentan en combinación con el isotig06880b y el isotig07270 respectivamente, mayor homología, en el primer caso con las secuencias denominadas fem-1 homolog CG6966; y en el segundo caso con las denominadas fem 1-like en especies relacionadas.

#### 3.3.2.1.1. Alineamiento general para las variantes de *feminizer*

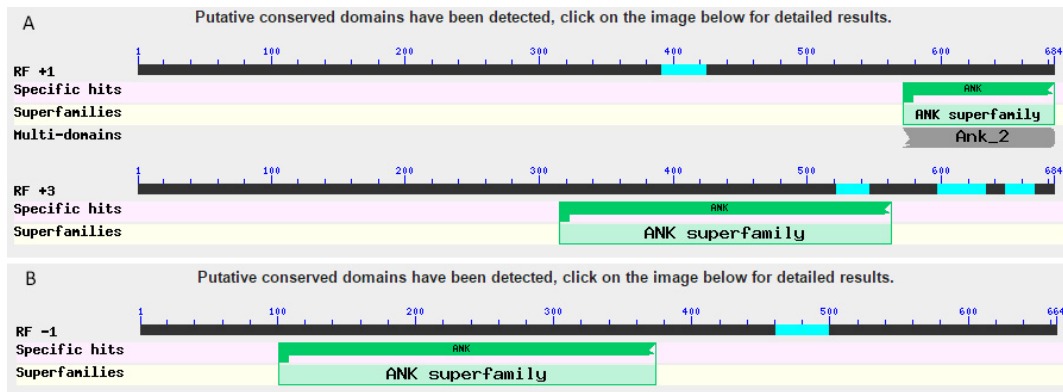
Al observar que los isotigs previamente mencionados podrían estar formando parte de secuencias nucleotídicas codificantes para variantes de la misma proteína, se llevó a cabo un alineamiento general de las secuencias proteicas que mostraron mayor similitud con los isotigs correspondientes a *D. longicaudata* por medio de BLASTx.

Es interesante notar que si bien las secuencias son diferentes existen zonas comunes a ambas variantes, la variante 1 correspondiente a *fem-1* homolog CG6966 y la variante 2, *fem-1* “homolog B”. El inicio de ambas variantes, al igual que la región central, presenta mayor variabilidad. Las regiones conservadas en las secuencias (resaltadas, Anexo 7) como era de esperarse se encuentran en las zonas de los dominios funcionales, desde la posición 110 a la posición (recuadros sólidos, Anexo 7).

##### -Variante 1: Isotig06680 e Isotig06680b

Los alineamientos por medio de BLASTx a partir del isotig06680 y 06680b arrojaron dos dominios conservados putativos vinculados con la función del gen *fem*. En este caso sobre uno de los fragmentos del isotig06680 se pudo localizar el dominio conservado ANK\_2, de la superfamilia de las Ankrinas (Pfam12796) (Anexo 7). Se puede observar que según el marco de lectura que

se considere se traducirá una u otra versión del dominio conservado. Al probar los diferentes marcos de lectura, al alinear con los marcos +1 y +3 se logra evidenciar la homología con los dominios, siendo mejor la cobertura de secuencia al utilizar el marco en +3 (Figura 27 A) (cobertura 71%, identidad de 98% y e-value  $9e^{-80}$ ). Al alinear el isotig06680b, se observó que también posee un dominio conservado de ANK, de la superfamilia de las Ankrinas (pfam13637) con una cobertura del 99%, identidad del 92% y e-value de  $3e^{-132}$  (Figura 27 B). Para este fragmento el marco de lectura adecuado es el -1.



**Figura 27.** Alineamiento en BLASTx de las secuencias correspondientes al Isotig06880 (A) y 06880b (B).

Como puede observarse en la Figuras 27, el mayor porcentaje de homología en ambos casos es el mismo organismo y corresponde a la misma secuencia [XP\_015126732.1] perteneciente a *Diachasma alloeum* (Muesebeck) (Hymenoptera: Braconidae), un pequeño parasitoide del gusano de la manzana *Rhagoletis pomonella* (Walsh) (Diptera: Tephritidae). El hecho de que ambos fragmentos presentaran homología con la misma secuencia permitió alinearlos sobre la misma y así diseñar cebadores específicos a fin de recuperar la porción central faltante del transcripto (Anexo 7). Al observar el resto de las secuencias que presentan homología con los isotigs utilizados para construir la variante 1, se evidencia que también tienen un grado de homología con la variante denominada *feminizer 1-like*, si bien es menor esta región se corresponde con el dominio funcional conservado correspondiente a la familia de las Ank\_2.

A fin de tener un panorama comparativo entre las secuencias correspondientes a la variante 1 se alinearon las secuencias protéicas de *D. alloeum* y *F. arisanus* [XP\_011315450.1], especies cercanas filogenéticamente a *D. longicaudata*; *N. vitripennis* [XP\_008207912.1] y *A. mellifera* [XP\_006564184.1].



Al observar el alineamiento se ve que para *D. longicaudata* la secuencia proteica contiene un fragmento inicial que no está presente en la secuencia correspondiente a *D. alloeum*. También puede verse que los dos dominios conservados que posee la proteína se encuentran presentes en *D. longicaudata*.

A partir de los cebadores diseñados y utilizando ADNc como template se logró amplificar por PCR de punto final el fragmento central correspondiente al gen *fem* que no se había logrado secuenciar previamente. Con las secuencias alineadas se ensamblaron o confirmaron las diferentes porciones del transcripto, verificando la homología de secuencia (Anexo 8).

La longitud del fragmento que se logró ensamblar para esta variante de *fem* fue de 2068bp y se logró identificar sobre ella dos dominios conservados de la familia de las ankrinas: Ank\_2 y Ank\_4 según Pfam (pfam12796 y pfam13637, respectivamente).

Sobre este ARNm construido a partir de la secuenciación de los fragmentos obtenidos por PCR, se diseñaron los cebadores específicos para llevar a cabo las RT-qPCR. Se consideró, entre otras cosas, para la elección del mejor par de cebadores que estos flanquearan los segmentos más conservados de la secuencia.

-Variante 2: Isotig06220 e Isotig07202, *fem-1* homolog

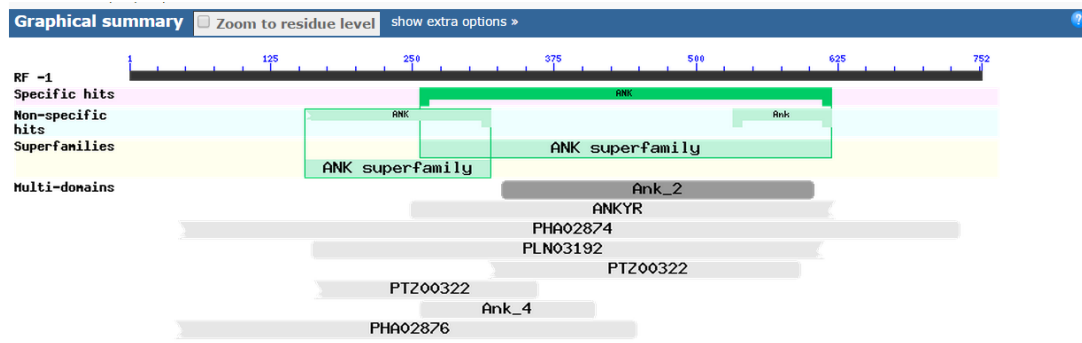
Los alineamientos a partir del isotig07202 y el isotig06260 mostraron nuevamente homología con el dominio conservado de las ankrinas, Pfam12796 Ank\_2. El isotig07202 mostró homología de secuencia para la traducción de la secuencia en el marco de lectura +3. Sin embargo no se alineó con dominios conservados (cobertura 20%; identidad del 98% y e-value de  $4e^{-19}$ ). En el caso del isotig06260 se observó homología de secuencia con la traducción en el marco de lectura en -1 (cobertura 99%; identidad del 88% y e-value de  $2e^{-159}$ ) para el dominio funcional que la identifica como *fem-1 homolog* (Figura 28). Ambos isotigs presentaron el mayor grado de homología y la mejor cobertura con la misma secuencia: *fem 1-homolog B* de *Fopius arisanus* (Sonan) (Hymenoptera: Braconidae), [XP\_011314979.1] (Anexo 8). *F. arisanus*, especie parasitoide de la mosca de la fruta utilizado como controlador biológico (Manoukis et al., 2011).

En este caso al alinear las secuencias de nucleótidos para ambos isotigs sobre la secuencia de *F. arisanus* [XP\_011314979.1] sobre la que mostraron el mayor grado de homología se observa que ambas se encuentran en los extremos de la secuencia dejando el sector central del transcripto sin cubrir. A fin de completar la secuencia del transcripto se amplificaron por PCR de punto final y se secuenciaron diferentes fragmentos (ver cebadores en materiales y métodos).

La secuencia construida para *D. longicaudata* ensamblando los resultados obtenidos por pirosecuenciación 454 y PCR de punto final logra cubrir la extensión de la secuencia de *F. arisanus* con la que mayor grado de similitud mostró, comprendiendo un total de 2654 nucleótidos. Los 174 nucleótidos iniciales de *F. arisanus* no se cubren con la secuencia de *D. longicaudata*, que a su vez posee una región final de mayor longitud.

A partir de las amplificaciones por PCR de punto final se logró identificar diferencias entre las secuencias provenientes de individuos hembras y machos. Si se observa el alineamiento en primer lugar, puede verse una inserción de una C en la posición 1036 para el caso de las hembras. A continuación a partir de la posición 1047 y hasta la posición 1330 en donde vuelve a observarse mayor similitud en la secuencia, se presenta la mayor diferencia (recuadro, Anexo 7). Sin embargo, algunas posiciones se encuentran conservadas incluso en la secuencia de *F. arisanus* (resaltadas en amarillo, Anexo 7). La posición 1425 marca el comienzo de una región variable que alcanza hasta la posición 1538. En este punto comienza para los machos, finalizando la posición 1611 en una serie de T, nuevamente una región diferencial.

Finalmente, a partir de la posición 1906 se comienza una región conservada para machos y hembras que se extiende hasta la posición 2363. A continuación se observa que la secuencia correspondiente a las hembras continúa hasta la posición 2828.



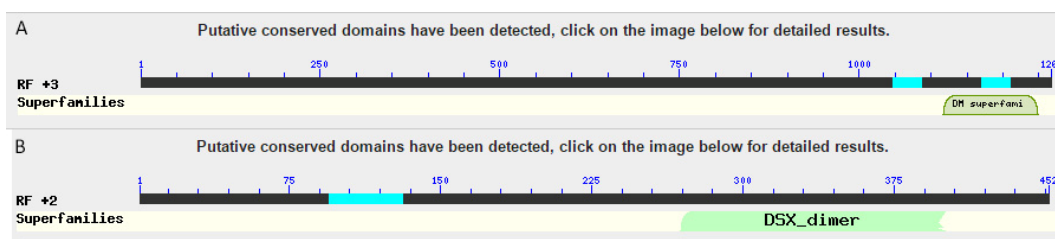
**Figura 28.** Alineamiento de los isotigs07270 y 06202 por medio de BLASTx con la base de datos del NCBI. Se observa para e isotig06202 que en el marco de lectura -1 alinea con la superfamilia de las ankrinas.

### 3.3.2.2. Doublesex: Isotig08283 e Isotigs00774/5

Los alineamientos de este grupo de secuencias en la base de datos del NCBI por medio de BLASTx mostraron homología a las secuencias previamente caracterizadas como *dsx* (Figura 32, Anexo 9). Las más altas homologías para estas secuencias fueron *F. arisanus*, proteína X1/X2[XP\_011307439.1] y *Microplitis demolitor* (Hymenoptera: Braconidae) isoformas X1/X2 [XM\_008556338.1]. En este caso se analizaron los isotigs 00774 y 00775 que presentaron homología de secuencia y se alinearon con el dominio conservado de unión al ADN del extremo N-terminal de la proteína. Para el isotig00774 el marco de lectura adecuado fue +3 con una cobertura 25%; identidad del 70% y e-value de  $1e^{-10}$ , mientras que para el isotig00775 el marco de lectura utilizado fue +2 con una cobertura del 25%; una identidad del 70% y un e-value de  $3e^{-9}$  (Figura 29 A, Anexo 9). Para el caso del isotig08283 en el marco de lectura +2 puede verse en la Figura 29 B que se alinea con el dominio conservado de dimerización de DSX (cobertura 88%; identidad; 70% y e-value  $2e^{-56}$ ).

En el alineamiento realizado entre las proteínas que presentaron el mayor grado de homología y los isotigs obtenidos en los marcos de traducción correspondientes, puede observarse que si bien las secuencias protéicas varían, en las zonas en las que se ubican los dominios funcionales de la proteína DSX podemos observar para las secuencias correspondientes a *D. longicaudata* la presencia de los aminoácidos conservados en estas posiciones. Esta situación se repite tanto para el dominio de unión al ADN presente en las posiciones del alineamiento 27 a 80, como para el dominio de dimerización del extremo C-terminal de la proteína, presente en las posiciones 180 a 239 del alineamiento (Anexo 9). En el extremo C-terminal se pueden observar en el recuadro azul los aminoácidos que se corresponden con las diferentes isoformas de la proteína, tanto para las secuencias de *F. arisanus* y *M. demolitor*, como para *D. longicaudata*. En el último caso, se hallaron tres secuencias diferentes para la proteína.

En el caso de este transcripto se logró cubrir la secuencia casi en su totalidad por medio del alineamiento de secuencias obtenidas por medio del experimento de pirosecuenciación, pudiendo incorporarse a la secuencia, además de los ya mencionados isotigs algunos singletons que al traducirlos en el marco correcto alinearon con las secuencias DSX de *F. arisanus* y *M. demolitor* permitiendo construir de manera más completa la secuencia para *D. longicaudata* (Anexo 9).



**Figura 29.** Alineamiento de los isotigs 08283 (A) y 0774 en BLASTx.

Al alinear las variantes de las secuencias nucleotídicas correspondientes a *dsx* para las especies con las que mayor homología presentaron los isotigs de *D. longicaudata*, se observa que las variantes de “splicing” del ARNm para las especies se encuentran en el centro de la secuencia y sobre el final, presentando diferentes longitudes. Adicionalmente se observa que las secuencias tienen segmentos de diferentes longitudes al inicio que no presentan un alto grado de similitud (Anexo 10).

Al incorporar las secuencias nucleotídicas obtenidas para *D. longicaudata*, tanto por pirosecuenciación como por secuenciación de productos de PCR de punto final (ver materiales y métodos), observamos un patrón similar al previamente descrito para las secuencias de las especies relacionadas. Se tomó de cada especie la secuencia de mayor homología y sobre ella se alinearon las secuencias correspondientes a *D. longicaudata*. Como puede observarse, la región inicial de los isotigs 00774 y 00775 está conservada pero presenta un bajo nivel de homología con las dos especies con las que se la está comparando, así como estas muestran que sucede lo mismo entre ellas. La zona de mayor homología entre secuencias se da en la región entre los 1227-1381 nucleótidos para *D. longicaudata* con 1082-1236 para *M. demolitor*. En la región central del transcripto es en donde mayores diferencias se observan, incluso entre las secuencias correspondientes a *D. longicaudata*. En esta zona se observa en primer término un corrimiento de tres pares de bases para una de las variantes (posiciones 1381 a 1383 del alineamiento); en segundo término la región del 1523 al 1616 presenta dos “gaps” en una de las variantes que terminará en esta posición. Se observan varios “gaps” de corta extensión en la región entre los 1667 a 1705 nucleótidos y uno de mayor extensión entre los 1820 y 1844 nucleótidos. Alrededor de la posición 1898 se ve el final de otra de las variantes encontradas y la secuencia de mayor longitud contiene 2209 pb (Anexo 10).

### 3.4. Análisis de expresión diferencial entre sexos: genes de determinación del sexo

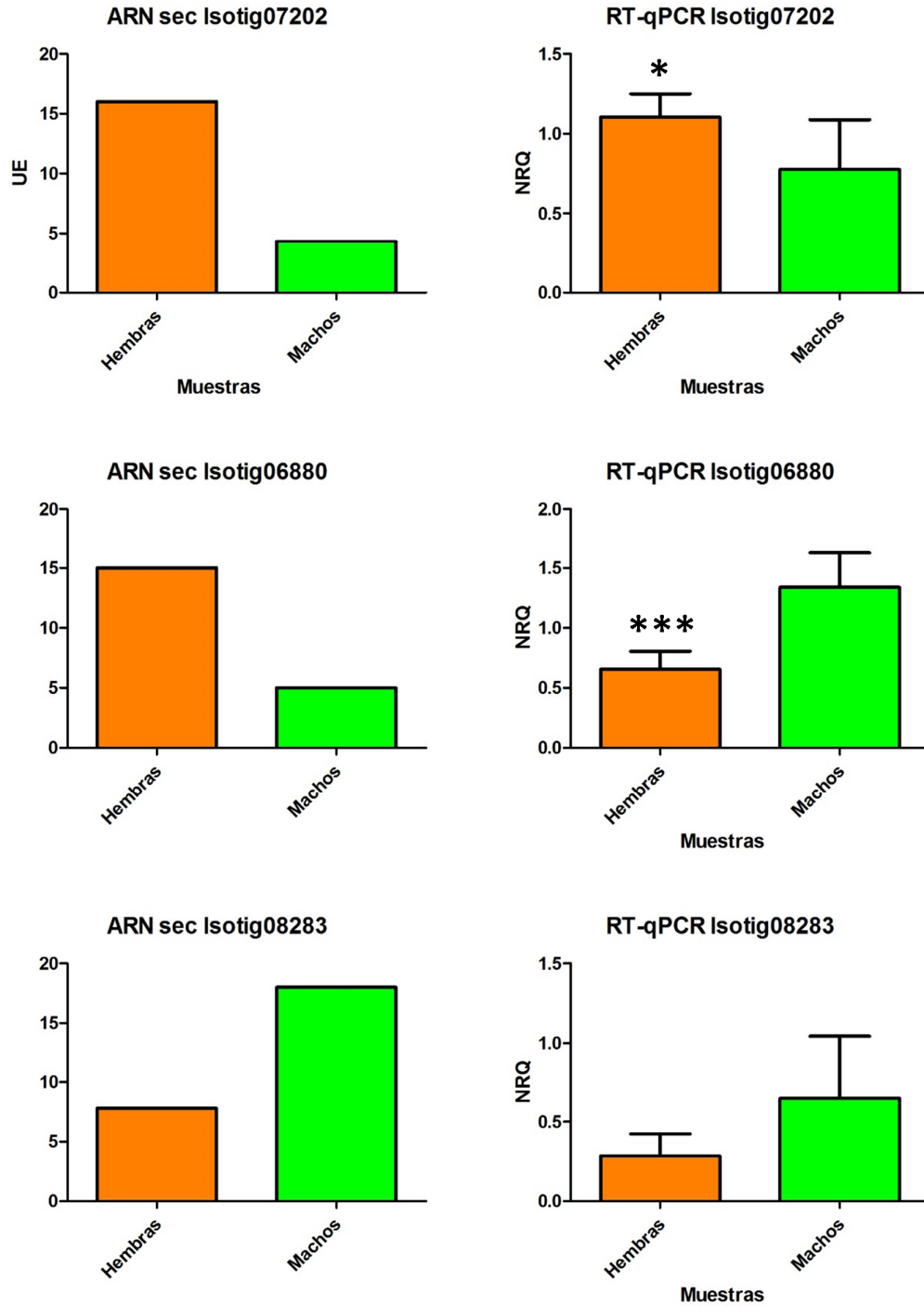
#### 3.4.1. Expresión diferencial en adultos

A fin de evaluar si los transcritos identificados por homología como variantes de *fem* y *dsx* se encontraban expresados diferencialmente entre hembras y machos adultos se analizaron tanto los resultados obtenidos por secuenciación NGS como los que se obtuvieron por medio de RT-qPCR. Ninguno de los transcritos que se identificaron alcanzó los valores de corte como para afirmar que se encontraban diferencialmente expresados por medio de los resultados de transcriptómica. Con la finalidad de evaluar la existencia de la diferencia de expresión mediante RT-qPCR se diseñaron cebadores específicos para los transcritos y se evaluaron la eficiencia de amplificación y el rango dinámico en cada caso (Tabla 11).

En principio, los resultados de pirosecuenciación muestran que para ambas variantes del mensajero *fem* habría una sobreexpresión en hembras adultas (Figura 30). Sin embargo, como se mencionó no alcanzaron a superar el valor de corte, por lo que las tendencias observadas tuvieron que ser evaluadas en todos los casos por RT-qPCR. En el caso del isotig06880, o variante 1, la tendencia observada por medio de secuenciación de ARN fue la sobreexpresión en hembras adultas. Al evaluar esta diferencia por medio de RT-qPCR, se revierte la tendencia observada. El resultado arrojado por esta técnica indica que este ARNm se encuentra sobreexpresado en machos con un R asociado de 1,8 ( $p=0,0001$ ) (Figura 33). Estos patrones de expresión opuestos detectados para el isotig06880 pueden deberse a que se contaba con un bajo recuento de lecturas provistas por la secuenciación de ARN. La RT-qPCR conlleva un análisis estadístico robusto y es por ello una técnica mucho más confiable.

Para el isotig07202, denominado anteriormente como variante 2, se observó por medio de la pirosecuenciación de ARN una tendencia que marcaba una sobreexpresión en hembras. Si bien no alcanzaba el valor de corte fijado anteriormente para evidenciar una diferencia de expresión concreta, podía estimarse una sobreexpresión en hembras adultas de *D. longicaudata*. Esta tendencia fue evaluada por medio de RT-qPCR y fue confirmada, mostrando un R de 1,2 veces mayor expresión en hembras adultas ( $p=0,02$ ) (Figura 30). Estos patrones de expresión opuestos encontrados para las variantes de *fem* podrían responder a sus funciones específicas en los diferentes sexos. Este aspecto requiere una caracterización funcional más profunda.

Para el isotig08283, homólogo a *dsx*, no se detectó una diferencia de expresión significativa entre sexos por medio de ninguna de las dos técnicas aplicadas (Figura 30). La diferencia por medio de RT-qPCR arrojó un valor marginalmente significativo ( $p=0,09$ ) luego del análisis estadístico. Si se observan las tendencias, habría mayor expresión en machos. Ya sea por baja cantidad de lecturas en secuenciación de ARN o por un error relativamente grande en los valores normalizados de Cq para RT-qPCR no se pudieron obtener diferencias significativas en este caso. Esto podría deberse a que la regulación a cargo del efector *dsx* se da fundamentalmente por “splicing” alternativo del mensajero en lugar de tratarse de una diferencia de expresión.



**Figura 30.** Comparación entre los resultados de expresión obtenidos por medio de NGS y RT-qPCR para los GI homólogos a *fem* y *dsx*. En la columna de la izquierda se observan los resultados de NGS y en la columna de la derecha se observan los resultados obtenidos por RT-qPCR para los mismos transcritos.

### 3.4.2. Expresión diferencial a lo largo del desarrollo

Para el estudio y evaluación de los GI que se han identificado como homólogos a *fem* y *dsx*, involucrados en las cascadas de determinación del sexo, se llevó a cabo un análisis a lo largo del desarrollo de individuos de *D. longicaudata*. Los genes involucrados en estas vías regulatorias son objeto de varios estudios incluso en la actualidad y es fundamental evaluar los cambios en sus patrones de expresión durante el desarrollo.

#### 3.4.2.1. Obtención de muestras y establecimiento de modificaciones a los protocolos

##### 3.4.2.1.1. Relación de masa y cantidad de individuos vs estadio del desarrollo:

Para llevar a cabo las extracciones de ARN a lo largo del desarrollo se ajustó, principalmente para los estadios iniciales, la cantidad de individuos utilizados como material de partida a fin de alcanzar una cantidad aceptable de ácido nucleico. En la Tabla 12 se listan las cantidades de individuos utilizadas para conformar cada réplica biológica. Asimismo, se muestran en la Tabla 12 a modo de ejemplo, resultados de la extracción de ARN y las medidas de calidad e integridad de algunas muestras. Las RT se llevaron a cabo según el protocolo establecido en la sección 1.1.3. Teniendo en cuenta que *D. longicaudata* es, como se mencionó anteriormente, haplodiploide y que es imposible distinguir entre hembras y machos en estadios inmaduros al menos hasta el estadio de pupa, se diseñó el experimento de manera que pudiera distinguirse el efecto de la presencia de machos en la muestras en contraste con muestras compuestas por ambos sexos. Con este propósito, se compararon dos grupos de individuos. El primero compuesto por individuos provenientes de hembras vírgenes, que estará compuesto exclusivamente por machos; y el segundo, compuesto por individuos provenientes de la cría que se conforma tanto de machos como de hembras.

**Tabla 12.** Muestras de *D. longicaudata* en el desarrollo. Cantidad de individuos utilizados por réplica biológica, parámetros de calidad. Se muestra un valor representativo para cada estadio.

Estadio	Cantidad de individuos	Concentración (ug/ul)	Relaciones de absorbancia	
			230/260	280/260
Huevo	60	650,3	1,70	1,85
Larva 1	30	982,1	1,82	1,75
Larva 3	6	3452,9	1,80	1,93
Pupa	10	2567,4	1,94	2,03



Adulto	15	2713,5	2,13	2,00
--------	----	--------	------	------

### 3.4.2.2. Cuantificación a lo largo del desarrollo: validación de los GR

Al momento de cuantificar es requisito contar con genes de expresión estable en todas las condiciones a ser analizadas. El caso particular del desarrollo plantea un desafío al momento de encontrar candidatos que cumplan estas condiciones. La expresión génica a lo largo del desarrollo es altamente variable ya que la diferenciación celular es dependiente de los patrones de expresión combinados de una gran cantidad de genes, motivo por el cual es dificultoso hallar aquellos que se mantengan constantes.

Al contar con los dos GR utilizados anteriormente, se decidió evaluar la posibilidad de utilizarlos a fin de poder contrastar todos los estadios en las condiciones ensayadas. Los valores obtenidos para los parámetros CV y M se muestran en la Tabla 13. Si bien la variación interna reflejada en el CV es mayor que para las muestras de adultos únicamente, es lógico que así sea debido a que a lo largo del desarrollo los patrones de expresión de la  $\beta$  actina varían de un estadio al otro en algunas especies (Sindelka et al., 2006). En este caso ambos GR analizados presentaron valores aceptables para ser utilizados como GR a lo largo del desarrollo y entre grupos de las dos condiciones analizadas.

**Tabla 13.** Validación de GR para estadios del desarrollo.

Gen	CV	M
<i><math>\beta</math>-actina</i>	25,15	0,20
<i>Factor de elongación 1 <math>\alpha</math></i>	24,04	0,21

### 3.4.2.3. Patrones de expresión para los GI a lo largo del desarrollo

Con el objetivo de evaluar el comportamiento de los GI identificados previamente en la sección 3.3. se midió el patrón de expresión de los mismos a lo largo del desarrollo. Como se mencionó anteriormente, es imposible distinguir entre machos y hembras en estadios tempranos del desarrollo de *D. longicaudata*. Por este motivo y con la finalidad de distinguir posibles diferencias ligadas al sexo, los grupos comparados en esta sección se compusieron, por un lado de individuos de la cría de *D. longicaudata* que serán denominados grupo control y por otro lado de individuos provenientes de hembras vírgenes. El primer grupo se compuso de una mezcla de

machos y hembras que corresponde a la proporción sexual característica de la cría. El segundo grupo se conformó de individuos machos exclusivamente.

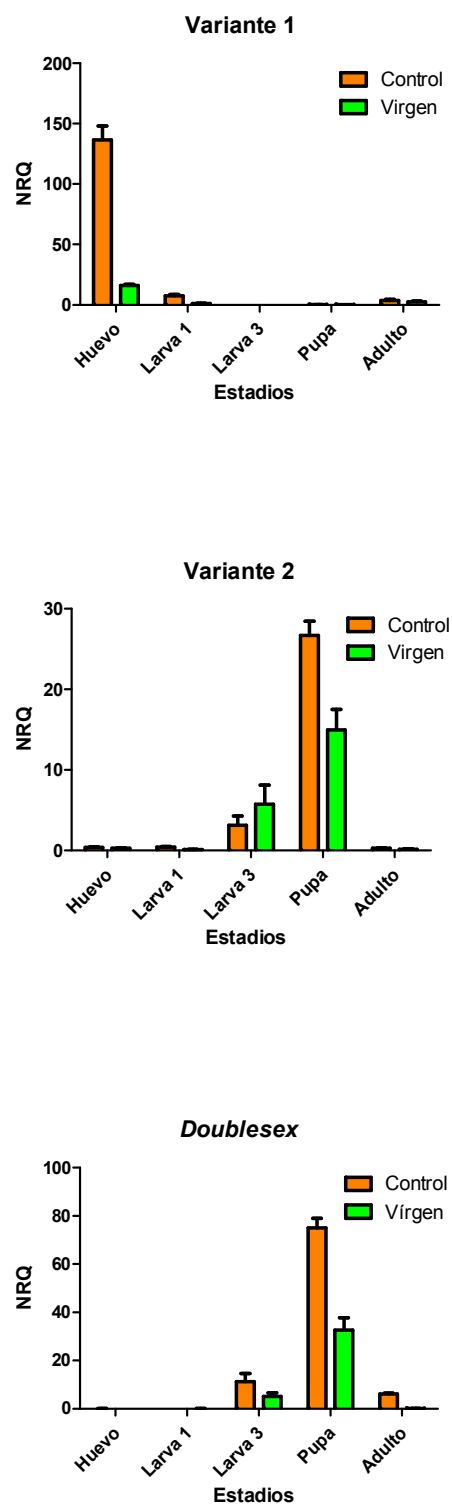
En la figura 31 se observan los gráficos correspondientes a los perfiles de expresión de los GI. La información obtenida para construir los perfiles de expresión, tanto de las dos variantes identificadas como *fem* así como la secuencia construida para *dsx* debe analizarse e interpretarse con cuidado.

En el caso de la variante 1, correspondiente a los isotigs06880 y 06880b, el perfil presenta un pico de expresión en el estadio inicial al observar la expresión en el tiempo. Al comparar entre muestras control y muestras vírgenes puede observarse que el NRQ es muy superior para las muestras control. El análisis estadístico por medio de ANOVA de dos factores muestra que existen efectos de interacción estadísticamente significativos entre los factores analizados. Por este motivo, no es posible una interpretación clara y en conjunto de este grupo de resultados (Anexo 11). Se procedió a continuación a llevar a cabo el análisis de los factores por separado a través de ANOVA de un factor. Para los diferentes estadios del desarrollo, el único que presenta diferencias significativas con el resto es el Huevo (Anexo 11). Esto es coherente con lo observado en la figura 31 A. Para las comparaciones entre control y virgen existen diferencias significativas en todos los estadios comparados. En el caso de los Huevos se sobreexpresa el transcripto en el grupo virgen alrededor de 10X ( $p < 0,0001$ ). Para las larvas 1, la sobreexpresión en las muestras virgen es de alrededor de 7X ( $p = 0,001$ ). En las larvas 3 la sobreexpresión sigue dándose en las muestras vírgenes (aproximadamente 1,5X,  $p = 0,0004$ ). En pupas sigue manteniéndose una diferencia de expresión significativa (2X,  $p = 0,002$ ) y adultos se revierte esta patrón, mostrando una sobreexpresión en los controles (2X,  $p = 0,01$ ).

Al analizar el comportamiento de la segunda variante, correspondiente a los isotigs 07202 y 06260, es evidente que en el estadio de pupa se da un pico de expresión del transcripto (Figura 31 B). Al analizar los datos de NRQ (ANOVA de dos factores) para esta variante se observó que las contribuciones a la variación observada las aportan tanto los factores analizados como la interacción entre ellos. Es decir, las interacciones son estadísticamente significativas por lo que no puede hacerse una interpretación de estos resultados en conjunto. Al tomar los factores por separado y analizarlos por medio de ANOVA de un factor, se obtiene que para los diferentes estadios del desarrollo la variación en el NRQ es significativa, siendo las pupas el único estadio que presenta diferencias significativas (Anexo 11). Al analizar el segundo factor, la condición, no se observan diferencias significativas no en Huevos ni en Larvas 3. Las Larvas 1 presentan sobreexpresión de alrededor de 4 veces en el grupo control ( $p = 0,004$ ). También se observan diferencias de expresión significativas en las Pupas ( $p = 0,003$ ) y en los adultos ( $p = 0,012$ ); sobreexpresándose el transcripto alrededor de 2 veces en los controles para ambos

casos. En este caso, las tendencias que se observaron en el gráfico concuerdan con el análisis estadístico.

En el caso de los isotigs 08382 y 00774/5, homólogos a *dsx*, puede verse en el gráfico que los niveles de transcripto aumentan hasta alcanzar un pico en el estadio de pupa (Figura 31 C). Cuando se analizan los resultados del ANOVA de dos factores, este arroja que las interacciones entre los dos factores son significativas por lo que se requiere analizarlos individualmente. Al analizar el factor Estadio del desarrollo por medio de ANOVA de un factor se ve que existen diferencias significativas entre las muestras, atribuyéndose al estadio de pupa ( $p < 0,0001$ , Test *Post Hoc* de Bonferroni). Tomando la Condición como factor de análisis, los ANOVA de un factor muestran que existen diferencias significativas entre condiciones para Larvas 3 (2X,  $p = 0,04$ ), Pupas (2X,  $p = 0,0003$ ) y Adultos (60X,  $p < 0,0001$ ). En todos los casos la sobreexpresión se da en el grupo control.



**Figura 31.** Expresión diferencial a lo largo del desarrollo. Para datos estadísticos ver el anexo 11.

# DISCUSIÓN GENERAL

## DISCUSION

*Diachasmimorpha longicaudata* es un insecto endoparasitoide de especies de moscas dípteras y es actualmente considerado uno de los principales agentes de control biológico de estas plagas. Particularmente es utilizado en los programas de liberaciones aumentativas para el control de especies de moscas de la fruta de los géneros *Anastrepha* y *Ceratitis* en las zonas tropicales y subtropicales de América (Suarez et al., 2014). *D. longicaudata* presenta una alta flexibilidad adaptativa a las diferentes condiciones ambientales en las que ha sido liberado (Purcell, 1998; Montoya et al., 2007; Vargas et al., 2012); se adapta a la multiplicación en condiciones artificiales de cría sobre diversos hospedadores como *C. capitata*, *A. fraterculus* y *A. ludens* (Ovruski et al., 2003b) y presenta gran habilidad para hallar a su hospedador (López et al., 1999; Sivinski et al., 2000; García-Medel et al., 2007). Es escaso el conocimiento que se tiene acerca de la genética y metabolismo de esta especie. En el grupo de trabajo se ha avanzado en la caracterización citogenética de *D. longicaudata* (Carabajal Paladino, 2011) y en aspectos genéticos referentes a su sistema de determinación sexual (Carabajal Paladino et al., 2015). Actualmente se conoce que esta especie presenta una determinación del sexo mediada por alelos complementarios con más de dos loci involucrados. Con esta información de base, se planteó avanzar en el conocimiento de las bases moleculares de la determinación sexual en *D. longicaudata* así como generar herramientas moleculares que permitieran evaluar la variabilidad genética de sus poblaciones. Los conocimientos generados en esta tesis resultan de importancia para el conocimiento y manipulación de las tasas sexuales de este insecto bajo cría artificial, de manera de favorecer la producción de hembras, sexo que permite implementar el control biológico en el campo. Esta información contribuirá a mejorar el rendimiento y la calidad de los insectos producidos ayudando a fortalecer la implementación de técnicas de control biológico amigables con el ambiente en nuestro país por medio del uso de parasitoides.

Con la finalidad de conocer los genes intervinientes en la cascada de determinación del sexo en *D. longicaudata* se utilizaron distintas estrategias que involucraban la obtención de muestras de ARN y ADNc de alta calidad. Mediante la optimización de los protocolos se logró establecer una rutina de trabajo para la obtención de ARN de alta calidad evaluando los diferentes parámetros involucrados. Si bien se contaba con experiencia previa en el grupo de trabajo en la especie relacionada *A. mellifera*, este último presenta mayor tamaño que el parasitoide y posibilita la extracción de material genético a partir de un solo individuo. En contraste, la especie en estudio, *D. longicaudata* es un insecto pequeño y resultó dificultoso obtener ARN de la concentración y calidad adecuadas a partir de un solo parasitoide para poder llevar a cabo correctamente tanto

la RT como obtener la cantidad y calidad de material necesario para la construcción de bibliotecas de ADNc que serán utilizadas para la obtención del transcriptoma por pirosecuenciación.

Si bien para las diferentes aplicaciones de la técnica de extracción de ARN, como Pirosecuenciación 454 y RT, se pueden introducir ligeras variaciones en el protocolo para obtener mayor cantidad de ácido nucleico o una mejor calidad del material, pueden hacerse algunas generalizaciones. En primer lugar, se observa que en general los mejores resultados para esta especie se encontraron al extraer ARN a partir de grupos de individuos. Cabe destacar que al incrementar el número de individuos se obtuvo una mejor representación de los transcritos presentes, y asimismo, una mayor contaminación de la muestra con ADN. Esto puede deberse a que en una extracción de ARN utilizando un número elevado de individuos es más factible que, por encontrarse en gran concentración, aparezca ADN como contaminante, generando problemas particularmente al momento de llevar a cabo experimentos de cuantificación por RT-qPCR. En segundo lugar puede decirse que el volumen de Trizol recomendado por el fabricante permitía una extracción de calidad aceptable, aun así, pudo observarse que al incrementar ese volumen los parámetros de calidad mejoraban sustancialmente (Figura 5, Tabla 1). Es clara la razón por la cual esto sucede, ya que de esta sustancia depende la estabilización del ARN y purificación eficiente del resto de los componentes del organismo, particularmente los lípidos y la quitina que se encuentran en gran proporción. Teniendo en cuenta dichos parámetros en este trabajo se determinó el número óptimo de individuos necesario y el volumen de Trizol adecuado para cada estadio del desarrollo utilizado, así como también para individuos adultos teniendo en cuenta la relación entre la cantidad de Trizol y la cantidad de individuos y la buena representación de los transcritos.

Las reacciones de Transcripción Reversa (RT) llevadas a cabo para obtener el ADN copia (ADNc) son un punto clave del procesamiento de las muestras. Es de vital importancia obtener un ADNc íntegro y de calidad para lograr amplificar por PCR fragmentos de ARN mensajero (ARNm) de cierta longitud y la calidad de la misma no se visualiza sino hasta el final de la aplicación para la cual fue necesaria la obtención del ADNc. En el caso de las reacciones de RT utilizando ARN de *D. longicaudata*, las condiciones no sufrieron grandes modificaciones a las recomendadas por el fabricante. Asimismo, se pudo observar que la calidad del ADNc no difiere al utilizar cualquiera de los dos tipos de cebadores (random primers u oligodt). Puede decirse entonces que, si la extracción de ARN ha sido llevada a cabo siguiendo el protocolo optimizado, este paso brinda mayor seguridad para la obtención de ADNc de calidad, aun después de tiempos prolongados de almacenamiento.

Siguiendo los esquemas clásicos para hallar genes candidatos en organismos no modelo, en este trabajo el primer intento por elucidar el mecanismo regulatorio encargado de la determinación del sexo en *D. longicaudata* consistió en la identificación de regiones de ARNm del parasitoide por homología de secuencia con especies relacionadas. Los resultados para el gen *csd* mostraron que las secuencias expresadas amplificadas a partir de ADNc, poseían una secuencia idéntica a las previamente identificadas para ADN genómico (Lanzavecchia et al., 2008). Del mismo modo que ocurrió con dichas secuencias previamente obtenidas, las nuevas secuencias no presentaron similitud de secuencia a nivel de secuencia de nucleótidos y aminoácidos con secuencias anotadas en GenBank.

Para el caso del gen *dsx*, no fue posible obtener secuencias a partir de las regiones amplificadas sobre ADNc para ninguna de las combinaciones de cebadores que se utilizaron. En el caso de *dsx*, la presencia de producto de amplificación únicamente en las hembras podría representar fragmentos de un producto de “splicing” alternativo correspondiente a la variante femenina de este transcripto. Al no poder obtener una secuencia de nucleótidos de calidad de estos fragmentos no se puede afirmar que lo fueran. A pesar de estas dificultades, al observar las regiones sobre las que estos cebadores fueron diseñados y compararlas con los resultados de transcriptómica, podría tratarse de la región nucleotídica codificante para la región variable entre machos y hembras identificada en especies relacionadas (Oliveira et al., 2009).

Al observar que los senderos sinuosos del enfoque clásico de búsqueda de genes candidatos por homología de secuencia con especies relacionadas no arrojaba resultados claros y aprovechando el advenimiento de las técnicas de secuenciación masiva, la siguiente etapa de estudio consistió en la búsqueda de genes específicos expresados a partir de la secuenciación del transcriptoma de *D. longicaudata*.

Las técnicas de secuenciación masiva o de próxima generación (NGS, del inglés “Next Generation Sequencing”) permiten la obtención y análisis de grandes cantidades de información de forma rápida. Representan uno de los medios relativamente modernos para la caracterización rápida de organismos no modelo, como *D. longicaudata*, brindando múltiples aplicaciones. A partir de esta tecnología fue posible generar el primer transcriptoma de *D. longicaudata* para su caracterización general, prestando especial atención a los posibles patrones sexo-específicos y permitiendo la generación de marcadores moleculares. De esa manera, a partir de las tres muestras de ARN (adulto hembra, adulto macho y larva 3 macho, ver detalles en sección materiales y métodos) se generaron tres bibliotecas de ADNc que fueron secuenciadas y ensambladas. Al observar los rendimientos de las tres bibliotecas puede evidenciarse que a



pesar de haber partido de cantidades similares de material se obtuvieron diferentes cantidades de lecturas. Al analizar otros ensayos llevados a cabo en condiciones similares, las bibliotecas de ADNc no tienen igual rendimiento (Nishimura et al., 2012). En el caso particular de *D. longicaudata*, machos y hembras poseen, debido a su naturaleza haplodiploide, diferente cantidad de cromosomas (Carabajal Paladino, 2011). Al momento de analizar la aparición de transcriptos ensamblados a partir de las diferentes bibliotecas, esta diferencia en la cantidad de lecturas no representa una diferencia en la posibilidad de encontrar representados a los transcriptos. Es decir, de algún modo se compensan estas diferencias entre sexos al momento de la expresión génica y regulación. La compensación de dosis para evitar un desbalance de expresión entre los sexos es vista habitualmente como una estrategia para evitar potenciales efectos dañinos causados por un desbalance de expresión a gran escala (Marin et al., 2000). Cuando la comparación se lleva a cabo entre adultos y larvas, se puede justificar la mayor cantidad de lecturas en la larva, dado que en este estadio del desarrollo previo a la metamorfosis los insectos se encuentran en una etapa de alta actividad metabólica regulatoria a causa de los grandes cambios a los que se enfrenta el organismo en este momento (Barmina et al., 2005; Casper and Doren, 2006). Estas variaciones son inherentes a la muestra analizada, ya que al evaluar el resto de los indicadores de calidad para cada biblioteca se ven valores similares, señalando que en esta etapa no se detectó ninguna diferencia entre una muestra y otra (Figuras 15 a 17, Tablas 5 a 7).

El 99,8% de las lecturas pudieron utilizarse para ensamblar el transcriptoma luego de evaluar la calidad de las mismas y filtrar adaptadores y cebadores. Si bien el ensayo fue acotado a una porción de la placa, el porcentaje de lecturas que pudieron utilizarse fue mayor al obtenido en otros organismos analizados por medio de la misma tecnología (Schwartz et al., 2010; Nishimura et al., 2012). El ensamblado arrojó 51686 unigenes, 8483 isotigs que se agruparon en 7295 isogrupos representando potenciales variantes de “splicing” de ARNm y 43203 singletons.

En relación a trabajos llevados a cabo en *C. vestalis* (Nishimura et al., 2012), parasitoide de características similares a *D. longicaudata*, puede observarse que las lecturas se obtuvieron por medio de dos metodologías de secuenciación: Illumina® y pirosecuenciación 454 (Roche). Las lecturas obtenidas por medio de la segunda técnica fueron de mucha menor cantidad que las obtenidas para *D. longicaudata*, pero al ensamblar se obtuvo un número mucho mayor de contigs en relación a los obtenidos en este trabajo. Esto puede vincularse con el hecho de que las longitudes promedio de lectura que se obtuvieron para *C. vestalis* fueron de alrededor de 500bp y en el caso de *D. longicaudata* fueron de alrededor de 400bp. Si bien en comparación con otros trabajos realizados en artrópodos (Zagrobelyny et al., 2009; Pauchet et al., 2010) esta

longitud promedio es superior se puede decir en líneas generales que el rendimiento de este tipo de técnicas es altamente sensible en relación a la calidad del ARN procesado. Otro parasitoide recientemente secuenciado, *F. arisanus*, arrojó un número de isotigs similar al obtenido para *D. longicaudata*, si bien la secuenciación se llevó a cabo por otra técnica de NGS (Illumina®) en la que se obtuvieron 8307 unigenes, unos cien menos que los obtenidos en este trabajo. Es llamativo que en *C. vestalis* el número de isotigs sea de alrededor de 17.000, no debe perderse de vista que la longitud de lectura fue algo más larga que en los otros dos casos. *F. arisanus* y *D. longicaudata* comparten muchas características en términos de su biología, no es ilógico pensar que ambos transcriptomas están representando un panorama acerca de la porción de genoma expresada en este tipo de insectos. En el caso de *C. vestalis*, tiene otro tipo de hospedador y esto podría explicar en parte la proporción de genes expresados que se encontraron para esta especie.

Puede decirse entonces que se obtuvo una cantidad de lecturas acorde a las esperadas para un ensayo de esta naturaleza y en las condiciones en las que fue llevado a cabo, permitiendo de esta forma continuar con la identificación de las secuencias que lograron ensamblarse.

Como era de esperarse, de los 8483 isotigs y 43203 singletons, la mayor parte pudieron anotarse por medio de BLASTx (83% y 47% respectivamente) como similares a especies dentro del orden Hymenoptera, mayoritariamente con grupos de hormigas y *A. mellifera* (Figura 20) en proporciones similares. Esto es congruente con la información que se encontraba disponible al momento de llevar a cabo la anotación para *D. longicaudata* ya que los transcriptomas de una serie de parasitoides fueron secuenciados recientemente (i5k workspace). Las muestras no se encontraban contaminadas ni siquiera con secuencias del hospedador sobre el que fueron criados (*C. capitata*). Al comparar nuevamente con las especies recientemente secuenciadas se obtienen porcentajes similares de transcriptos anotados, para *F. arisanus* se anotaron el 85% de los unigenes (Calla et al., 2015).

En relación a los péptidos predichos para ambas especies, *D. longicaudata* y *F. arisanus*, se obtuvo alrededor de la mitad en el primer caso debido a las diferentes técnicas aplicadas para obtener las secuencias en ambos casos. Pero a diferencia de *F. arisanus*, las longitudes promedio en *D. longicaudata* fueron aproximadamente un 33% más largas.

Los transcriptos obtenidos para *D. longicaudata* a través de los genomas y transcriptomas de las dos especies modelo de himenópteros *A. mellifera* y *N. vitripennis* muestra una distribución uniforme. A su vez, pueden observarse transcriptos presentes a lo largo de todos los

cromosomas en ambos casos. Esto indica que se obtuvo un transcriptoma con una buena representación y cobertura, sin sesgos hacia un grupo particular de transcriptos.

Como una de las múltiples aplicaciones de la obtención del transcriptoma de *D. longicaudata* fue de interés la obtención *a priori* e *in silico* de marcadores moleculares. En la actualidad es escasa la información disponible acerca de la genética del complejo de especies crípticas *D. longicaudata* y los microsatélites descubiertos tienen potencial aplicación en la caracterización de la variabilidad genética tanto en poblaciones silvestres como de laboratorio (Jenkins et al., 2011; Julsirikul et al., 2014). Esta especie se ha establecido bajo condiciones de cría artificial en varios países de América para ser usada en programas integrados de control biológico contra especies de tefrítidos, moscas plaga de los frutos de importancia económica (Montoya et al., 2012; Vargas et al., 2012). En este momento, el parasitoide está siendo criado en Argentina con el mismo objetivo (Suarez et al., 2014). La presencia de poblaciones silvestres de *D. longicaudata* en nuestro país que ha sido reportada previamente por Schliserman et al., (Ovruski et al., 2003b) y Oroño y Ovruski (Oroño and Ovruski, 2007), se originó a partir de liberaciones a campo llevadas a cabo con anterioridad en el marco de programas de control biológico en nuestro país. Los marcadores aquí presentados podrían contribuir a la descripción de estas poblaciones a nivel genético así como podrían ser útiles en el desarrollo de herramientas de diagnóstico molecular que permitan analizar la calidad genética de poblaciones de laboratorio y la producción a escala de biofábrica de esta especie (determinación de niveles de endogamia, genotipos deletéreos, etc.). Adicionalmente, el uso de marcadores moleculares permite la construcción de mapas de ligamiento genético que permitan el análisis de características fenotípicas y comportamentales que han sido exploradas en otras especies de insectos (Holloway et al., 2000; Hunt and Page, 1995) pero que representan un desafío en el caso de la genética y genómica de *D. longicaudata*. Esto podría revelar información acerca de las principales vías metabólicas y su regulación así como la comprensión de las bases genéticas de características complejas como la determinación del sexo, la parasitización, la localización y preferencia por un hospedador, etc. Toda esta información es de gran valor para ser aplicada al mejoramiento de los protocolos de cría masiva y liberación de este agente de control biológico (Montoya et al., 2012; Ovruski and Schliserman, 2012; Vargas et al., 2012).

La anotación funcional de secuencias para *D. longicaudata* permitió asignar un total de 40241 términos de ontología génica (GO) a 16930 transcriptos. La mayoría de estos términos se asignaron a procesos elementales metabólicos y en segundo lugar a procesos involucrados en respuestas a estímulos ambientales. En relación a los términos que se encuentran en mayor

proporción dentro de las actividades metabólicas se encuentran las de unión, catálisis y transporte.

Ahondando en la identificación de transcriptos por ontología génica se encontraron cuatro secuencias vinculadas a la determinación del sexo y otras diez asociadas a la diferenciación sexual y procesos reproductivos vinculados a la determinación del sexo. En cuanto a los transcriptos asociados a determinación sexual, no pudieron identificarse las secuencias, solamente se les atribuyó la función tanto para la línea somática como para la germinal. En este caso solo fue posible ubicar marcadores moleculares sobre una sola secuencia y en el caso de las siete secuencias adicionales fue posible identificar marcadores moleculares en todas ellas que además presentaron secuencias flanqueantes lo suficientemente largas como para permitir el diseño de cebadores específicos. Un análisis más profundo de estas secuencias por medio de RACE PCR permitirá obtener mayor número de marcadores potencialmente asociados a la vía regulatoria de determinación del sexo.

A partir de una comparación inicial entre los transcriptomas para hembra y macho adultos, se evidenció cualitativamente la presencia de transcriptos que se expresaban diferencialmente entre los sexos. De entre los transcriptos que presentaban diferencias de expresión elevadas entre los sexos, se seleccionaron tres de estos que se expresaban con una notable diferencia y se evaluaron por medio de RT-qPCR, confirmando estas diferencias. Ninguna de estas secuencias estuvo entre las identificadas como involucradas en determinación del sexo, ni por homología de secuencias ni por ontología génica. Estos tres transcriptos se identificaron como Isotig01415; Isotig02512 e Isotig03526, homólogos funcionales a las siguientes proteínas: Proteínas de unión a Odorantes; Esfingomielina diesterasa y Myrosinquinasa (Tabla 11). Si bien para estos transcriptos el resultado fue el mismo por medio de ambas técnicas no se puede hacer extensivo el resultado a todas las secuencias evaluadas, por lo que fue necesario establecer un valor de corte a partir del cual es necesario evaluar los transcriptos por medio de RT-qPCR. Este valor de corte se definió en base a dos parámetros. En primer lugar, una expresión del al menos 10 veces más entre condiciones y en segundo lugar, las secuencias a ser comparadas debían aparecer al menos tres veces en las muestras comparadas.

Al hacer un análisis comparativo enfocando la atención sobre los patrones de expresión correspondientes a cada sexo en individuos adultos se hallaron 98 transcriptos que se expresaban diferencialmente, es decir el 1,2% del transcriptoma es sexo-específico basándonos en el valor de corte previamente definido. Esto indica una gran estabilidad en la expresión génica en individuos adultos. No hay que perder de vista que al analizar los genes que presentan

patrones de expresión sexo-específica en muestras compuestas de mezclas de tejidos (individuos completos) los resultados observados pueden responder a dos causas. En primer lugar, las diferencias podrían deberse a una mayor concentración total de ARNm por sobreexpresión en un tejido específico; es decir, una concentración total mayor por contribución de un tejido único del individuo. En segundo lugar, la causa de las diferencias podría ser la expresión en múltiples tejidos, que también resultará en una diferencia total que ahora se deberá a la expresión en todo el individuo en conjunto. Al analizar los tejidos individualmente es muy probable que la cantidad de genes con expresión diferencial asociada al sexo aumente ya que si las diferencias de expresión no son elevadas pueden enmascarse al tomar como referencia la totalidad del individuo (Ellegren and Parsch, 2007). La mayor cantidad de transcritos mostraba sobreexpresión en hembras, particularmente para un grupo de isotigs que pertenecían al mismo isogrupo, es decir un locus genómico con variantes de transcritos diferentes. Al hacer un análisis más profundo de estas secuencias, se las encontró representadas en especies relacionadas de himenópteros como *N. vitripennis*, *F. arisanus*, *A. mellifera*, etc. Estas secuencias se dividían en dos grupos en *D. longicaudata*, una de las variantes contenía una delección de tres bases y algunos cambios de bases puntuales. Es interesante destacar que estas secuencias no tienen función asignada, ni en la base de datos de *D. longicaudata* ni en GeneBank en relación a las mencionadas especies de himenópteros. Es conocido que los genes con patrones de expresión sexo-específicos evolucionan a un ritmo mayor que el resto del genoma, especialmente los relacionados al sexo masculino, y que tienen una elevada tasa de variación de aminoácidos entre una especie y otra filogenéticamente cercanas (Manu et al., 2013). Esta gran divergencia propicia que no sea posible identificar fácilmente ortólogos entre especies relacionadas.

A partir de estos patrones de expresión sexo-específica obtenidos en individuos adultos surge la incógnita de buscar entre estos aquellas secuencias que por homología se identificaron como genes involucrados en la cascada de determinación del sexo. En primer término se retomó la identificación de secuencias del gen *csd* que se había intentado caracterizar inicialmente. Al alinear las secuencias obtenidas para el gen *csd* por medio de PCR de punto final con la base de datos de *D. longicaudata* generada por transcriptómica, tanto la región variable como la conservada del gen *csd*, presentaban similitud con secuencias (isotigs) diferentes, anotadas con funciones diferentes provenientes. En el mismo sentido, al analizar el fragmento de mayor tamaño que se obtenía al intentar amplificar la potencial región central del transcripto *csd*, se pudo interpretar que la amplificación era inespecífica o una quimera producto de la reacción de transcripción reversa ya que al intentar amplificar ese mismo fragmento sobre ADN nunca

fue posible (inicialmente se consideraba que esto se debía a la presencia de intrones). De todas maneras, como se mencionó anteriormente, estos resultados requieren de análisis de secuencia más profundos para arribar a conclusiones más consistentes.

Al buscar en la base de datos de *D. longicaudata* secuencias que presentaran homología con genes involucrados en las vías de determinación del sexo, se encontraron siete secuencias, cuatro de ellas mostraban homología con diferentes variantes del gen *feminizer* y tres con el efector *doublesex* (Tabla 11). A partir de estos datos, se profundizó la caracterización de estas secuencias obteniendo que las cuatro secuencias en realidad alineaban con dos variantes de *feminizer*. Los isotigs 07202 y 06260 alinearon sobre las secuencias homólogas a la proteína “feminizer 1 like” en los extremos de la misma. El centro del transcripto faltante que no pudo secuenciarse por NGS, se recuperó a partir de secuenciación por la metodología de Sanger obteniendo fragmentos amplificados por PCR de punto final a partir de ADNc (Anexo 7). Para el caso de los isotigs 06880 y 06880b sucedió algo similar, reconstruyéndose de igual modo el centro del transcripto que mostró similitud de secuencia con la proteína *feminizerGC* (Anexo 8). En el caso de los isotigs 08283 y 00774/00775 que mostraron homología con *dsx*, se confirmó la secuencia obtenida y se obtuvieron patrones diferentes de “splicing” para machos y hembras adultos (Anexo 9). A partir de la secuenciación y “reconstrucción” de secuencias para estos dos componentes claves en el módulo central de la cascada de determinación del sexo, pudieron identificarse variantes asociadas al sexo aunque es necesaria una caracterización funcional de las secuencias para confirmar su rol regulatorio.

Si bien es conocido que la regulación de estos transcriptos ocurre por “splicing” alternativo se hallaron resultados de expresión diferencial asociados al sexo, particularmente para dos variantes del gen *feminizer* que fueron identificadas. En el primer caso el Isotig07202/Isotig06260 fue identificado como homólogo a la proteína *feminizer 1* (Tabla 11) y ambas técnicas arrojaron una sobreexpresión en hembras adultas, si bien por pirosecuenciación solo se evidenció una tendencia. En el segundo caso, para el Isotig06880 homólogo a la proteína *feminizerGC* (Tabla 11), se revierte la tendencia observada por pirosecuenciación sobreexpresando significativamente en machos adultos. Estos resultados sugieren que si bien ambos transcriptos se encuentran presentes en los dos sexos, aun en individuos adultos se expresan diferencialmente sugiriendo algún rol regulatorio aun en estadio maduro. La diferencia marginalmente significativa que presentó el efector *dsx* puede deberse a que en este caso la regulación podría, al igual que sucede en especies relacionadas (Cho et al., 2007; Verhulst et al., 2010), se da por “splicing” alternativo del transcripto con presencia de variantes del mismo en individuos adultos de ambos sexos.

Para ahondar en el comportamiento de los genes *fem* y *dsx* y con la intención de esclarecer el cómo y el cuándo estos genes comienzan a ejercer su influencia, surge la necesidad de evaluar la expresión a lo largo del desarrollo. Para esto, fue necesario definir y optimizar genes de referencia que pudieran utilizarse en esta especie, tanto en individuos adultos como en el desarrollo para permitir la cuantificación de transcritos para *D. longicaudata*. Se identificaron a partir de bibliografía genes que ya habían sido comúnmente utilizados en himenópteros y otros órdenes y se evaluaron con resultados positivos la  $\beta$  *actina* y el *factor de elongación 1  $\alpha$*  (Kubista et al., 2006). Así como para la mayoría de las especies ambos genes mostraron un rango dinámico acorde con buena eficiencia de amplificación. Del mismo modo se logró optimizar las reacciones de RT-qPCR para los genes de interés evaluados en todos los casos. Se logró estandarizar el protocolo para *D. longicaudata* e identificar genes de referencia aplicables a diferentes condiciones.

En cuanto a los genes de interés evaluados a lo largo del desarrollo, pueden observarse patrones complementarios entre el isotig06880, correspondiente a la variante 1, y la variante 2 de *fem* en conjunto con el isotig08283, correspondiente al gen *dsx*. Esto es congruente con la regulación propuesta en otras especies para este sistema. El pico inicial de la variante 1 de *fem* podría provenir de la madre, como es sabido en esta instancia estos picos de expresión se corresponden más con el hecho de que la madre transfiere ARNm al huevoque debido a sobreexpresión en el mismo. Es decir, esto indicaría en *D. longicaudata* la presencia temprana de un regulador *fem/tra* que además de presentar variantes de “splicing” modula la posterior acción del efector *dsx*.

# CONCLUSIONES GENERALES



## Conclusiones

- Se consiguió la obtención de protocolos específicos de extracción de ARN, RT y qPCR puestos a punto para *D. longicaudata*.

-Se obtuvo el primer transcriptoma para *D. longicaudata*. La calidad de las secuencias obtenidas fue buena y se logró una buena distribución de transcritos sin sesgos, cubriendo la mayoría de los cromosomas de especies filogenéticamente relacionadas con representantes de transcritos del parasitoide.

- A partir de la pirosecuenciación del transcriptoma fue posible obtener una batería de marcadores moleculares que se encuentran ubicados sobre secuencias expresadas. Se identificaron y se ponen a disposición 2775 marcadores microsatélites para estudios de variabilidad genética en poblaciones naturales, introducidas y de laboratorio de *D. longicaudata*.

-El análisis del transcriptoma permitió anotar por GO el 37,4% de los transcritos totales (8451 isotigs correspondientes a variantes de genes). Se identificaron 11 secuencias potencialmente vinculadas a la determinación del sexo en la especie 4 de ellas sin función conocida.

- Se hallaron 98 transcritos con una diferencia de expresión significativa por medio de pirosecuenciación 454. El 1,2% del transcriptoma es sexo específico. Los patrones de expresión son muy similares entre sexos para este organismo. Se validaron los resultados para los patrones de expresión asociada al sexo RT-qPCR.

- Se describieron por primera vez las secuencias nucleotídicas de los genes *fem* y *dsx*. Se identificaron variantes de “splicing” alternativo para ambos genes.

-El análisis de los patrones de expresión de las variantes del transcrita *fem* mostró que se expresan diferencialmente en adultos. Para *dsx* se hallaron diferencias marginalmente significativas.

- La expresión de los genes *fem* y *dsx* a lo largo del desarrollo muestra patrones que podrían responder a una regulación coordinada en el tiempo del regulador *feminizer* y el efector *doublesex*.

## Conclusión general

Este trabajo de tesis brinda los primeros conocimientos genómicos generales en *D. longicaudata* y sienta los fundamentos claves para el estudio de las bases moleculares de la determinación del sexo en esta especie haplodiploide, mediante la identificación de genes y el análisis de los perfiles de expresión en estadios tempranos del desarrollo. Estos genes constituyen potenciales herramientas biotecnológicas para la manipulación de la tasa sexual en la producción masiva de este parasitoide, que en conjunto con las herramientas moleculares (microsatelites) aplicables al estudio de la variabilidad genética de poblaciones naturales del parasitoide y la calidad genética en condiciones de cría artificial constituyen dos pilares fundamentales para los estudios genéticos en esta especie aplicados a la utilización de *D. longicaudata* como agente de control biológico en el campo.

# MATERIALES Y METODOS

## MATERIALES Y MÉTODOS

### 1. Insectos:

Los individuos de *Diachasmimorpha longicaudata* utilizados provinieron de la cría experimental mantenida en el Laboratorio de Insectos de Importancia Económica “Ewald A. Favret” en el INTA Hurlingham (Buenos Aires, Argentina). La cepa utilizada para fundar la colonia fue importada desde México a Hurlingham en 1988, e introducida al INTA Castelar en el año 2001 (SENASA, Expte. n° 14054/98). Esta especie se mantuvo en condiciones de cría de nivel experimental utilizando como hospedador *Ceratitis capitata* en un ambiente controlado con una temperatura promedio de 25°C, 85% de humedad relativa y un fotoperíodo 18:6 (horas luz: horas oscuridad).

Para mantener la cría de laboratorio, las hembras adultas (sexualmente maduras) fueron sometidas a exposiciones de larvas en estadio 2-3 de *C. capitata* en dieta artificial (Terán, 1977). Estas exposiciones se realizaron en cajas de Petri plásticas (5cm de diámetro y 1 cm de profundidad) cubiertas con tela de algodón (“voile”) por un período de cuatro horas. Posteriormente, se transfirieron las larvas de *C. capitata* a bandejas plásticas con medio artificial fresco a fin de completar su desarrollo. Estas bandejas fueron colocadas dentro de otras que contenían una capa de vermiculita como sustrato de pupación. Tanto las larvas como las pupas de *C. capitata* (parasitizadas y no parasitizadas) se mantuvieron en las mismas condiciones ambientales descriptas previamente para los parasitoides adultos.

Para los ensayos de expresión, los estadios inmaduros del desarrollo del parasitoide se recuperaron de las larvas expuestas y los puparios parasitizados mediante disección bajo microscopio estereoscópico binocular Olympus SZ11.

### 2. Extracción de ADN:

El material de partida para las extracciones de ADN fue en cada caso individuos adultos, macho o hembra de quince días de edad. La extracción se llevó a cabo siguiendo el protocolo de Baruffi (Baruffi et al., 1995) modificado a fin de optimizar el procedimiento para insectos de pequeño tamaño. En cada caso la muestra se homogeneizó en Buffer de extracción (Solución I, Anexo soluciones).

### 3. Extracción de ARN:

Inicialmente se tomó como referencia el protocolo propuesto por el fabricante que indica la relación volumen de reactivo/masa inicial de tejido (1ml/100mg) y se tuvo en consideración que cada individuo adulto de *D. longicaudata* criado artificialmente sobre *C. capitata* poseía una

masa de alrededor de 3mg. El tamaño pequeño de los individuos adultos y el alto contenido de quitina dificultaron mucho la tarea de hacer extracciones de buena calidad y rendimiento a partir de un solo individuo. Para lograr la obtención de una masa de tejido suficiente para extraer 1mg de ARN recomendado por el fabricante como cantidad de material de partida para realizar la transcripción reversa, se llevaron a cabo pruebas con diferentes cantidades de parasitoides a fin de determinar el número de individuos que sería necesario procesar en conjunto como material de partida para cada muestra. Se ensayaron extracciones de ARN a partir de 5, 10, 15 y 30 individuos utilizando la proporción tejido/volumen de Trizol recomendada por el fabricante.

Cabe aclarar que para llevar a cabo el ensayo de secuenciación masiva de ARN por medio de la técnica de pirosecuenciación 454 (Roche) (servicio INDEAR, Rosario, Santa Fe, Argentina) fue necesario duplicar el material de partida por los requerimientos mínimos de la técnica de medida. Se exigía como mínimo una cantidad de 75ug totales de ARN de partida.

Luego de homogeneizar las muestras en Trizol se centrifugó el homogenato a 4°C. Luego de la separación de fases, en todos los casos se tomó un volumen fijo de fase acuosa (150ul) para continuar con la purificación. El ARN resultante se resuspendió en 50ul de agua DEPC. Para evaluar la calidad e integridad del ARN obtenido se midieron las muestras por medio de absorbancia en un Nanodrop 1000 (ND 2000) y se corrieron en geles de agarosa 1% m/V a 100mV por 15 minutos.

#### 4. Tratamiento con DNase I:

Se utilizó el kit DNase I de Promega siguiendo las indicaciones del fabricante. Para evaluar la calidad de los resultados se analizó el producto final a través de electroforesis en gel de agarosa de concentración 1% con una corriente constante de 30mA.

#### 5. Transcripción Reversa:

El material de partida fue en todos los casos 1ug de ARN de alta calidad. Se llevó a cabo la transcripción reversa utilizando el kit comercial ImProm-II Reverse Transcriptase (Promega) utilizando cebadores Random y Oligo(dT) inespecíficos a fin de obtener únicamente fragmentos de ARNm. El ciclo de amplificación para la obtención de cADN constó de tres pasos que se llevaron a cabo en un termociclador Veriti®96-Well Thermal Cycler (Applied Biosystems). En primera instancia, se desnaturalizó el ARN incubándolo junto a los cebadores por 5 minutos a 70°C y 5 minutos a 0°C. El paso de hibridización se realizó por 60 minutos a 43°C y finalmente la extensión fue de 15 minutos a 70°C. El ADNc resultante se almacenó a -20°C.

## 6. Selección de Genes de Interés (GI):

Para llevar a cabo las amplificaciones de la región *csd*, los cebadores habían sido diseñados previamente por el grupo de trabajo. Los cebadores utilizados para amplificar el gen *dsx* fueron diseñados a partir de alineamientos múltiples de secuencias aminioacídicas y sobre las regiones correspondientes a los dominios funcionales de las proteínas.

En este caso se buscaron ortólogos a partir de alineamiento de secuencias y diseño de cebadores degenerados a partir de estos alineamientos en regiones conservadas a nivel proteína y de ser posible a nivel AN también.

- Gen *csd* (“complementary sex determiner”): Para la caracterización de este gen en *D. longicaudata* se utilizaron cebadores obtenidos de bibliografía correspondientes a *A. mellifera* (Beye et al., 2003) diseñados sobre regiones conservadas y variables del transcripto. También se utilizó un par de cebadores diseñado previamente dentro del grupo de investigación sobre una secuencia de ADN de *D. longicaudata* (Lanzavecchia, 2007).
- Gen *dsx* (doble sexo): Para la caracterización de este gen se diseñaron cebadores degenerados a partir del alineamiento de proteínas Dsx y ARNm correspondientes a diferentes himenópteros y dípteros (Oliveira et al., 2009). Los cebadores elegidos se ubicaron sobre regiones conservadas entre especies en el ARNm y dentro de los dominios funcionales de las proteínas.

**Tabla 14.** Cebadores de PCR de punto final para los genes *csd* y *dsx*.

Cebadores de PCR			
Nombre	Secuencia (5'-3')	Región	Locus
CSD1F	ATCCTCAGCATCGCATTGCA	Variable	<i>csd</i> ( <i>A. mellifera</i> )
CSD1R	ACTTCTGTCGAGTAGCATCACTCA	Variable	
fwCSD	CTTGTTCCGGTATTWTCATAAA	Conservada	
revSL	TAATGAATGAAATATGAGATGACATT	Conservada	
consCSD	RTCATCTCATWTTTCATTATTCAAT	Variable	
dsxFC	NTGTCAGAATNACGGNAAGANGGTCCA	Conservada E2	<i>dsx</i> ( <i>N. vitripennis</i> )
dsxR10	GNCAATCGNTAACTNTGATGACGTA	Variable E5/E6	

## 7. PCR de punto final:

El ciclado utilizado fue un primer paso de desnaturalización de 5 minutos a 95°C; 35 pasos de amplificación que constaron de 30 segundos de desnaturalización a 95°C, 30 segundos de annealing a 60°C y 1 minuto de elongación a 72°C; y un paso final de elongación por 7 minutos a 72°C. Las amplificaciones se llevaron a cabo en un termociclador Verity® 96 Wells (Applied Biosystems). En todos los casos los resultados fueron evaluados en geles de agarosa del 1,5% de concentración que se sometieron a electroforesis por 45 minutos a 45mA constantes.

El molde utilizado fue tanto ADNdc como ADNc. En el cebador caso se utilizó una dilución 1/5 de la muestra y en el segundo caso una dilución 1/10. Las reacciones se llevaron a cabo con Taq Polimerasa (Kit T-Plus ADN Polimerasa, InbioHighway) y un ciclo de PCR estándar. En cada caso se incluyó un control negativo de reacción. Para todos los casos se utilizó un ciclado básico de PCR que constó de un paso de desnaturalización a 95°C por 5 minutos; un paso de annealing cuya temperatura fue definida por el cebador utilizado en cada caso y un paso final de extensión a 72°C por 7 minutos. Todas las amplificaciones se llevaron a cabo en un termociclador Verity® 96 Wells (Applied Biosystems).

8. Purificación: se utilizó el kit Wizard SV Gel & PCR Clean Up System (Promega) siguiendo las instrucciones del fabricante.

9. Secuenciación por el método de Sanger: todas las muestras purificadas fueron secuenciadas en el SIGYSA-CICVyA, INTA Castelar.

10. Análisis de secuencias: se utilizó Bioedit v7.2.0 (Ibis, Biosciences 2013). Se compararon las secuencias nucleotídicas obtenidas con la base de dato del NCBI por medio de BLASTn 2.2.26 (NCBI, Basic Local Alignment SearchTool; Altschul et al. 1997), así como la base de datos de Himenópteros (Hymenoptera Genome database [www.http://hymenopteragenome.org/](http://hymenopteragenome.org/)).

11. Pirosecuenciación 454:

Para obtener una cobertura representativa del transcriptoma de *D. longicaudata* se decidió utilizar una combinación de muestras tal que permitieran a la vez llevar a cabo un análisis general y representativo del transcriptoma de la especie así como de las diferencias particulares entre sexos. Con esta finalidad se incluyó en el análisis, además de una muestra de machos adultos y otra de hembras adultas, una muestra adicional de larvas de macho. Como se mencionó previamente, esta especie es endoparasítica y al desarrollarse en el interior de su hospedador dificulta mucho la recuperación de muestras en estadios tempranos del desarrollo. Adicionalmente, es imposible hacer una distinción de sexos previa al estadio del desarrollo de larva 3 de los mimos, imposibilitando la obtención de individuos de sexos diferentes. Para la

obtención de larvas 3 de machos se aprovechó la haplodiploidía de la especie (solamente se desarrollaran individuos machos a partir de huevos de hembras vírgenes y se obtendrá una muestra mixta a partir de huevos fecundados). Se utilizaron para la extracción de ARN total una muestra de larvas 3 de macho, una muestra de machos adultos y una muestra de hembras adultas. Luego del enriquecimiento en mARN y la obtención de las tres bibliotecas de cADN, la pirosecuenciación 454 fue llevada a cabo de manera independiente para las tres muestras en media placa de un pirosecuenciador 454 GS FLX Titanium (Roche).

Se extrajo ARN de tres muestras de individuos de *D. longicaudata*. Cada pool consistió de aproximadamente 90mg de tejido. En cada caso se correspondió con 30 hembras adultas, 30 machos adultos (ambas muestras se tomaron quince días después de la emergencia) y 5 larvas de machos del tercer estadio (ocho días posterior a la oviposición de hembras vírgenes). El pool de larvas se incluyó en el ensayo a fin de conseguir un set de datos más representativo. Aproximadamente 90ug de ARN de alta calidad fueron procesados en INDEAR (Instituto de Biotecnología de Rosario, Rosario, Argentina) para la obtención de bibliotecas de cADN enriquecidas en mARN por medio de columnas de poliT. Posteriormente se procedió a realizar la Pirosecuenciación 454. Las bibliotecas de cADN de *Diachasmimorpha longicaudata* fueron ensayadas en una corrida de media placa en el pirosecuenciador 454-GS-FLX (Roche).

#### 12. Ensamblado y anotación del transcriptoma:

En primer lugar, para cada muestra se llevó a cabo un análisis de calidad de las secuencias obtenidas a fin de filtrar aquellas que presentaran baja calidad o contuvieran secuencias de los adaptadores utilizados en la construcción de las bibliotecas de mARN (poner figuras de INDEAR en la parte de resultados).

Se utilizó el Software Newbler v2.6 Assembler (Roche, IN, US) para el ensamblado de todas las secuencias crudas obtenidas a partir de las tres bibliotecas en contigs (secuencias alineadas desde un solo transcripto), isotigs (variantes de alineamiento de los contigs) e isogroups (isotigs que representan un locus genómico). Todas aquellas secuencias que se identificaron como singletons (i.e., lecturas de secuencias que no pudieron ser ensambladas en ningún isotig) fueron analizadas por medio del algoritmo de agrupación CD-HIT-454 utilizando un valor de corte de identidad de secuencia del 90%. De este modo se eliminaron secuencias redundantes o duplicados artificiales. A fin de conseguir una identificación putativa de la identidad de las secuencias se realizó una búsqueda por medio del algoritmo BLASTx (valor de corte del valor de  $e \leq 10^{-10}$ ) comparando con la base de secuencias de proteínas nr del NCBI. La anotación y el mapeo se llevaron a cabo utilizando el software BLAST2GO (Conesa et al., 2005) que por medio



de una combinación de búsquedas de similitud y análisis estadístico permite identificar y clasificar las secuencias previamente ensambladas (Moriya et al., 2007). Dicho software se encarga de asignar términos de ontología (Gene Ontology terms, GO, <http://www.geneontology.org>) (Baranowski et al., 1993), Mapas KEGG (Kyoto Encyclopedia of Genes and Genomes, KASS) y un número de clasificación enzimática (EC number).

La anotación de GO se completó utilizando el software InterProScan full suite con los parámetros estándar (Zdobnov and Apweiler, 2001). Con la finalidad de obtener los marcos de lectura (ORF) más probables para cada secuencia ensamblada (tanto isotigs como singletons) se utilizó el programa TransDecoder (Haas et al., 2008) definiendo el análisis a partir de los parámetros estándar. Los mejores y más probables péptidos se utilizaron como set de datos de entrada para el programa InterProScan. InterProScan combina diferentes métodos de reconocimiento de firma de proteínas originarios de las bases de datos miembros de InterPro como recurso para búsqueda e identificación de las anotaciones InterPro y GO correspondientes.

Para la comparación de transcriptomas con otros organismos modelo se utilizó el paquete del software MUMMER, específicamente el paquete Nucmer (Kurtz et al., 2004), y la herramienta de visualización Circos (Krzywinski et al., 2009). Las bases de datos de organismos modelo utilizadas para llevar a cabo estas comparaciones fueron las siguientes: para *A. mellifera* se utilizó el Honey Bee Genome Sequencing Consortium assembly Amel\_4.5 and Official Gene Set OGSv3.2 ([www.beebase.org](http://www.beebase.org)) y para *N. vitripennis* se utilizó el Nasonia Base genome [Assembly Nvit 1.0](http://hymenopteragenome.org/nasonia/) (<http://hymenopteragenome.org/nasonia/>).

### 13. Obtención de potenciales marcadores moleculares: SSR y SNP

A fin de identificar potenciales marcadores moleculares se utilizó el software MISA [ref, <http://pgrc.ipk-gatersleben.de/misa/misa.html>]. El criterio utilizado para la selección de SSR se basó en un número mínimo de repeticiones de secuencia. Para los motivos dinucleótidos fue de cinco repeticiones, cuatro para los trinucleótidos y tres repeticiones para los motivos tetra, penta y hexanucleótidos.

Para el diseño de todos los pares de cebadores para la amplificación de los marcadores SSR en cada secuencia ensamblada, se utilizó un “in-house script” que utiliza el software PRIMER3 del paquete EMBOSS (<http://emboss.open-bio.org/wiki/Appdocs>).

### 14. Identificación de transcriptos diferencialmente expresados:

Se extrajo ARN total a partir de grupos de hembras y machos adultos de *D. longicaudata*, cada grupo consistió en 15 individuos. Se utilizó Trizol® (Invitrogen). El ARN total se resuspendió en 30 µl de agua tratada DPEC y la calidad/cantidad fue medida espectrofotométricamente por absorbancia 260/280 y 260/230, como por geles de agarosa desnaturalizantes al 1%. Alrededor de 1 µg de ARN total se usó como template para sintetizar ADNc de hebra simple usando el kit ImProm-II Reverse Transcriptase (Promega) y cebadores Oligo (dT) (Promega) según los protocolos detallados en resultados. El ADNc resultante fue diluido 1/10 y almacenado a -20°C para su posterior uso en RT-qPCR.

Para identificar contigs expresados diferencialmente, las lecturas de cada biblioteca se mapearon de nuevo sobre los contigs sobre los cuales se habían ensamblado previamente usando todas las bibliotecas en conjunto. Para evaluar la expresión diferencial entre sexos a partir de los resultados de Pirosecuenciación, el nivel relativo de expresión para un determinado transcripto se calculó normalizando la cantidad de lecturas con el largo de ese transcripto y dividiendo por el número de lecturas en esa biblioteca. Este valor luego fue multiplicado por la biblioteca de mayor tamaño (Larvas 3). A este resultado se le denomina Unidades de Expresión (UE). Para la identificación de transcriptos expresados diferencialmente, se restringió el análisis a los transcriptos que se encontraban presentes en las tres bibliotecas. Cualquier isotig con menos de tres lecturas se descartó debido a la gran variabilidad inherente a la estimación de la diferencia de expresión a partir de genes con baja cobertura. A fin de determinar un valor de corte a partir del cual considerar a un transcripto como diferencialmente expresado se fijó un valor de corte en al menos diez veces de diferencia entre las condiciones. El número de lecturas de hembras y machos se normalizó con la biblioteca de larvas y se representa en escala logarítmica.

#### 15. Selección de GI

La selección de GI se basó en dos criterios: 1) un set de secuencias que incluyeron a los isotigs 01415, 02512 and 03151, se eligió a fin de validar los resultados de expresión diferencial hembra/macho por medio de NGS; el segundo set se eligió a partir de la homología por medio de BLASTx y se seleccionaron los isotigs 06880 y 06880b, 07270 y 06260; homólogos al gen *fem* y los isotigs 08283, 00774/5, homólogos al gen *dsx*. Los cebadores utilizados fueron en todos los casos diseñados sobre los isotigs obtenidos, tanto para confirmar las secuencias como para completar los fragmentos faltantes.

*-Feminizer*

## Cebadores Variante 1

Nombre	Secuencia	Tm	Posición
fem1CG_fw2	AGCAGCCAGGATCAGTTGTC	60.04	400-422
fem1CG_rv2	TGTCGATTGGCTGCTTCGAT	60.11	619-641
fem1CG_fw3	GGTAGGACAACTGGTGAGCG	60.39	284-306
fem1CG_rv3	AGACAACTGATCCTGGCTGC	60.04	401-423
fem1CG_fw8	TGGTTGGGGAGGAAGGAAGA	60.10	1426-1448
fem1CG_rv8	TAAGGTCACGGTCGCAGTTG	60.32	1559-1581
fem1CG_fw10	GAGGCCTCAATAGCAGGCAG	60.53	1755-1777
fem1CG_rv9	CGATATGGGCACCAGCTCTC	60.53	1928-1950
fem1CG_fw11	GCAGCCAGGATCAGTTGTCT	60.04	620-642
fem1CG_rv13	CTTCCCTTTGCACTGGAGGT	59.89	1466-1488
Fem1CG_fw18	<u>TTCCCTGATGGCCAATGACT</u>	59	1309-1331
Fem1CG_rv23	TAGTAGTCATTGGCCATCAG	60.04	1312-1334

## Cebadores Variante 2

Nombre	Secuencia	Tm °C	Posición
fem1_fw1	CAACCACGCAACGCATACAA	60.04	216-238
fem1_rv1	ACAGAGTGCCTTCTCGTTGG	59.97	414-436
fem1_fw3	AATGACCCCACTGATCGCAG	60.11	531-553
fem1_rv3	ATGATGCACCGAGGAGTTCG	60.18	633-655
fem1_fw8	TTGTTCCAGGCAGCTAAGCA	59.89	1817-1839
fem1_rv8	GTCACGAGTATTGAGGGGCA	59.46	2033-2055
fem1_fw9	CTTCAGTGCATGGCTGCAAA	59.68	1468-1490
fem1_rv9	CCACATACCAAACAGCCACG	59.48	1653-1675
fem1_fw10	GGAGGTGATTGAGGCCTACG	59.89	615-637
fem1_rv13	CTTTTGAGCCATGCACTGA	59.68	1470-1492
fem1-R18	CGCTTTCTCATTGAGTGCCA	58.84	1221-1243
fem1-F22	GAGTCATTGGCCATCAGGGA	59.45	824-846
fem1-F23	GCCACTGCACAGTTGTTGATT	59.93	1403-1425

*-Doublesex*

Nombre	Secuencia	Tm °C	Posición
DSX-F1	GTCCCAACAACATCCACTCGT	60	1535-1557
DSX-R1	TCTGCCTTGGCATCCTTGAG	60	1677-1699
DSX-F3	ATGAGGTGGAACCACAACCA	60	1342-1364
DSX-R3	TGGCCTTCATCAGCAACA	60	1747-1769
DSX-F4	GTCAGTACCACAGCCAGCTC	60	1388-1410
DSX-R4	ATGTTGTGGGACTGCGGTAG	60	1524-1546
DSX-F6	CTGGACGTCAGGACTGCTTT	59.96	789-808

DSX-R6	TTGCAATGCCATCACCTCT	59.97	1228-1250
DSX-F5	TGCCTGTGTAGCGATACGTG	60	107-126
DSX-F8	AAGAAAGGGCGCTGACTGTGCG	60	703-722
DSX-R8	CTCTCTTCAACGCCGTTTGC	60	1247-1269
DSX_F9	ACTGGGGAGGTTGGACCTTA	60	524-543
DSX-R9	CCGATAACCCACGTGCAGAT	60	629-648
DSX-R10	TCCAACCTCCCCAGTTACCA	60	519-538
DSX-F7	ACGAAGCTCGAAGACGCATT	60.4	1699-1721

#### 16. PCR Cuantitativa:

##### -Diseño de cebadores específicos

A partir de las secuencias listadas en la Tabla 11 se diseñaron los cebadores con los que se llevaron a cabo las reacciones de RT-qPCR. Para cada caso se evaluó la Eficiencia de amplificación, el rango dinámico de la reacción y el ajuste lineal de las curvas de calibración (Kubista et al., 2006).

##### -Determinación de eficiencia de amplificación y curvas de calibración

Para cada par de cebadores diseñados se evaluó la eficiencia de reacción por medio de la amplificación de una serie de diluciones del producto de PCR de dichos cebadores purificado. Se utilizaron diluciones seriadas con un factor de 100 a fin de construir una regresión lineal que permita calcular la eficiencia de amplificación y la linealidad de los puntos. Se aceptan como válidas aquellas presente eficiencias de entre 80 y 110.

##### - qPCR:

Todos los GI se midieron por medio de RT-qPCR analysis usando IQ<sup>TM</sup>SYBR®Green Supermix (Bio-Rad). Se diseñaron los cebadores listados en el Anexo X usando la herramienta Primer-BLAST tool (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Los parámetros de ciclado fueron 95°C por 5 min seguidos de 40 ciclos de 95°C por 10 s y 60°C por 45 s finalizando con la construcción de una curva de desnaturalización del producto amplificado. La expresión relativa se analizó por medio del método de múltiples GR (Hellemans et al., 2007). El *Factor de Elongación 1-alfa* ( $ef1-\alpha$ ) y la  $\beta$  *actinadeD. longicaudata* fueron usados como GR internos, como se han usado en otros insectos (Lourenço et al., 2008; Rong et al., 2013). El análisis estadístico se llevó a cabo por medio del test de t de Student o por medio de ANOVA de dos factores seguido de contrastes *a posterior* por medio del test de Bonferroni según lo requiera el diseño de los experimentos.

# BIBLIOGRAFÍA

## Bibliografía

- Asplen, M., Whitfield, J., De Boer, J., Heimpel, G., 2009. Ancestral state reconstruction analysis of hymenopteran sex determination mechanisms. *J. Evol. Biol.* 22, 1762–1769. doi:10.1111/j.1420-9101.2009.01774.x
- Bachtrog, D., Mank, J.E., Peichel, C.L., Kirkpatrick, M., Otto, S.P., Ashman, T.-L., Hahn, M.W., Kitano, J., Mayrose, I., Ming, R., Perrin, N., Ross, L., Valenzuela, N., Vamosi, J.C., 2014. Sex determination: why so many ways of doing it? *PLoS Biol.* 12, e1001899. doi:10.1371/journal.pbio.1001899
- Badouin, H., Belkhir, K., Gregson, E., Galindo, J., Sundström, L., Martin, S.J., Butlin, R.K., Smadja, C.M., 2013. Transcriptome characterisation of the ant *Formica exsecta* with new insights into the evolution of desaturase genes in social hymenoptera. *PLoS One* 8, e68200. doi:10.1371/journal.pone.0068200
- Bai, X., Zhang, W., Orantes, L., Jun, T.-H., Mittapalli, O., Mian, M. a R., Michel, A.P., 2010. Combining next-generation sequencing strategies for rapid molecular resource development from an invasive aphid species, *Aphis glycines*. *PLoS One* 5, e11370. doi:10.1371/journal.pone.0011370
- Baranowski, R.M., Glenn, H., Sivinski, J., 1993. Biological control of the caribbean fruit fly (Diptera: Tephritidae). *Florida Entomol.* 76, 245–251.
- Barmina, O., Gonzalo, M., McIntyre, L.M., Kopp, A., 2005. Sex- and segment-specific modulation of gene expression profiles in *Drosophila*. *Dev. Biol.* 288, 528–44. doi:10.1016/j.ydbio.2005.09.052
- Baruffi, L., Damiani, G., Guglielmino, C., Bandi, C., Malacrida, A., Gasperi, G., 1995. Polymorphism within and between populations of *Ceratitis capitata*: comparison between RAPD and multilocus enzyme electrophoresis data. *Heredity (Edinb.)* 74, 425–437.
- Beldade, P., Rudd, S., Gruber, J., Long, A., 2006. A wing expressed sequence tag resource for *Bicyclus anynana* butterflies, an evo-devo model. *BMC Genomics* 7, 130.
- Bess, H., 1961. Fruit fly parasites and their activities in Hawaii. *Proc. Hawaiian Entomol. Soc.* 367–378.
- Beye, M., 2004. The dice of fate : the *csd* gene and how its allelic composition regulates sexual development in the honey bee , *Apis mellifera*. *BioEssays* 26, 1131–1139. doi:10.1002/bies.20098
- Beye, M., Hasselmann, M., Fondrk, M., Page, R., Omholt, S.W., 2003. The Gene *csd* Is the Primary Signal for Sexual Development in the Honeybee and Encodes an SR-Type Protein. *Cell* 114, 419–429.
- Bopp, D., Saccone, G., Beye, M., 2014. Sex determination in insects: variations on a common theme. *Sex Dev.* 8, 20–8. doi:10.1159/000356458

- Calla, B., Sim, S.B., Hall, B., Derego, T., Liang, G.H., Geib, S.M., 2015. Transcriptome of the egg parasitoid *Fopius arisanus* : an important biocontrol tool for Tephritid fruit fly suppression. *Gigascience* 4, 1–5. doi:10.1186/s13742-015-0075-4
- Cancino, J., Lopez, E., Aguilar, C., 1995. Liberaciones inundativas de parasitoides como método alternativo de control de *Ceratitis capitata* en fincas cafetaleras en el Soconusco, Chiapas, México, in: *Actas de La 1° Conferencia Internacional IFOAM Sobre Café Orgánico*, IFOAM/AMAE/Universidad Autónoma de Chapingo, México. pp. 51–53.
- Carabajal Paladino, L., 2011. Genética y citogenética de la determinación del sexo en *D. longicaudata* (Hymenoptera, Braconidae). Universidad Nacional de Buenos Aires, Facultad de Ciencias Exactas y Naturales.
- Carabajal Paladino, L., Muntaabski, I., Lanzavecchia, S., Le Bagousse-Pinguet, Y., Viscarret, M., Juri, M., Fueyo-Sánchez, L., Papeschi, A., Cladera, J., Bressa, M.J., 2015. Complementary Sex Determination in the Parasitic Wasp *Diachasmimorpha longicaudata*. *PLoS One* 10, e0119619. doi:10.1371/journal.pone.0119619
- Carabajal Paladino, L., Papeschi, A., Cladera, J., 2010. Immature stages of development in the parasitoid wasp , *Diachasmimorpha longicaudata*. *J. Insect Sci.* 10, 1–9. doi:10.1673/031.010.5601
- Carabajal Paladino, L., Papeschi, A., Lanzavecchia, S., Cladera, J., 2013. Cytogenetic characterization of *Diachasmimorpha longicaudata* (Hymenoptera : Braconidae ), a parasitoid wasp used as a biological control agent. *Eur. J. Entomol.* 110, 401–409.
- Casper, A., Doren, M. Van, 2006. The control of sexual identity in the *Drosophila* germline. *Development* 2791, 2783–2791. doi:10.1242/dev.02415
- Cho, S., Huang, Z.Y., Zhang, J., 2007. Sex-specific splicing of the honey bee doublesex gene reveals 300 million years of evolution at the bottom of the insect sex-determination pathway. *Genetics* 177, 1733–41. doi:10.1534/genetics.107.078980
- Clausen, C., Clancy, D., Chock, Q., 1965. Biological control of the oriental fruit fly (*Dacus dorsalis* Hendel) and other fruit flies in Hawaii. *Tech. Bull. United States Dep. Agric.* 322, 67–102.
- Conesa, A., Gotz, S., Garcia-Gomez, J., Terol, J., Talon, M., Robles, M., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., Robles, M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674–3676. doi:10.1093/bioinformatics/bti610
- Debach, P., Rosen, D., 1991. Biological control by natural enemies. Cambridge Univ. Press.
- DeBach, P., Rosen, D., 1991. The natural enemies, in: *Biological Control by Natural Enemies*. pp. 35–84.
- Delano-Frier, J.P., Aviles-Arnaut, H., Casarrubias-Castillo, K., Casique-Arroyo, G., Castrillon-Arbelaez, P. a, Herrera-Estrella, L., Massange-Sanchez, J., Martinez-Gallardo, N. a, Parra-Cota, F.I., Vargas-Ortiz, E., Estrada-Hernandez, M.G., 2011. Transcriptomic analysis of grain amaranth (*Amaranthus hypochondriacus*) using 454 pyrosequencing: comparison

- with *A. tuberculatus*, expression profiling in stems and in response to biotic and abiotic stress. *BMC Genomics* 12, 363. doi:10.1186/1471-2164-12-363
- Devescovi, F., 2015. Estudios sobre el superparasitismo de larvas de la mosca del Mediterráneo *Ceratitis capitata* (Diptera : Tephritidae) por el parasitoide *Diachasmimorpha longicaudata* (Hymenoptera : Braconidae). UBA.
- Ellegren, H., Parsch, J., 2007. The evolution of sex-biased genes and sex-biased gene expression. *Nat. Rev. Genet.* 8, 689–698. doi:doi:10.1038/nrg2167
- Eo, S.H., Doyle, J.M., Hale, M.C., Marra, N.J., Ruhl, J.D., Dewoody, J.A., 2011. Comparative transcriptomics and gene expression in larval tiger salamander (*Ambystoma tigrinum*) gill and lung tissues as revealed by pyrosequencing. *Gene* 492, 329–338. doi:10.1016/j.gene.2011.11.018
- Evrogen, 2015. RNA analysis on non-denaturing agarose gel electrophoresis [WWW Document]. Support. *Methods Nucleic Acid Res.* URL <http://www.evrogen.com/technologies/RNA-electrophoresis.shtml>
- Favila, M., Halffter, G., 1997. The use of indicator groups for measuring biodiversity as related to community structure and function. *Acta Zoológica Mex.* 72, 1–25.
- Fernández, F., Sharkey, M., 2006. Sistemática de los himenópteros de la Región Neotropical: Estado del conocimiento y perspectivas, in: *Introducción a Los Hymenoptera de La Región Neotropical*. Sociedad Colombiana de Entomología y Universidad Nacional de Colombia, Bogotá D.C., Bogotá, pp. 7–35.
- García-Medel, D., Sivinski, J., Díaz-Fleischer, F., Ramírez-Romero, R., Aluja, M., 2007. Foraging behavior by six fruit fly parasitoids (Hymenoptera: Braconidae) released as single- or multiple-species cohorts in field cages: influence of fruit location and host density. *Biol. Control* 43, 12–22.
- Gempe, T., Hasselmann, M., Schiøtt, M., Hause, G., Otte, M., Beye, M., 2009. Sex determination in honeybees: two separate mechanisms induce and maintain the female pathway. *PLoS Biol.* 7, e1000222. doi:10.1371/journal.pbio.1000222
- Gempe, T., Beye, M., 2011. Function and evolution of sex determination mechanisms, genes and pathways in insects. *Bioessays* 33, 52–60. doi:10.1002/bies.201000043
- Geuverink, E., Beukeboom, L.W., 2014. Phylogenetic distribution and evolutionary dynamics of the sex determination genes *doublesex* and *transformer* in insects. *Sex Dev.* 8, 38–49. doi:10.1159/000357056
- Gish, W., States, D., 1993. Identification of protein coding regions by database similarity search. *Nat. Genet.* 3, 266–272.
- González, P.I., Montoya, P., Perez-Iachaud, G., Cancino, J., Liedo, P., 2007. Superparasitism in mass reared *Diachasmimorpha longicaudata* (Ashmead) (Hymenoptera : Braconidae), a parasitoid of fruit flies (Diptera : Tephritidae). *Biol. Control* 40, 320–326. doi:10.1016/j.biocontrol.2006.11.009



- Goulet, H., Huber, J.T., 1993. Order Hymenoptera, Hymenoptera of the world: An identification guide to families.
- Greany, P., Ashley, T., Baranowski, R., Chambers, D., 1976. Rearing and life history studies on *Biosteres (Opius) longicaudatus* [Hym.: Braconidae]. *Entomophaga* 21, 207–215.
- Greany, P., Tumlinson, J., Chambers, D., Bousch, G., 1977. Chemically mediated host finding by *Biosteres (Opius) longicaudatus*, a parasitoid of Tephritid fruit fly larvae. *J. Chem. Ecol.* 3, 189–195.
- Guillén, D., Sánchez, R., 2007. Expansion of the National Fruit Fly Control Programme in Argentina, in: M. J. B. Vreysen, A.S.R. and J.H. (Ed.), *Area-Wide Control of Insect Pests*. pp. 653–660. doi:10.1007/978-1-4020-6059-5\_60
- Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., Wortman, J.R., 2008. Automated eukaryotic gene structure annotation using EvidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* 9, R7. doi:10.1186/gb-2008-9-1-r7
- Hagen, B., 1953. A premating period in certain species of the genus *Opius* (Hymenoptera: Braconidae). *Proc. Hawaiian Entomol. Soc.* 15, 115–116.
- Hanson, P., Gauld, I., 1995. *The Hymenoptera of Costa Rica*. Oxford Univ. Press 44, 136. doi:10.1002/mmnd.19970440203
- Harpur, B.A., Sobhani, M., Amro, Z., Zayed, A., 2012. A review of the consequences of complementary sex determination and diploid male production on mating failures in the Hymenoptera. *Entomol. Exp. Appl.* 146, 156–164. doi:10.1111/j.1570-7458.2012.01306.x
- Hasselmann, M., Gempe, T., Schiøtt, M., Nunes-silva, C., Otte, M., Beye, M., 2008. Evidence for the evolutionary nascence of a novel sex determination pathway in honeybees. *Nature* 454, 519–523. doi:10.1038/nature07052
- Hasselmann, M., Lechner, S., Schulte, C., Beye, M., 2010. Origin of a function by tandem gene duplication limits the evolutionary capability of its sister copy. *Proc. Natl. Acad. Sci. U. S. A.* 107, 13378–83. doi:10.1073/pnas.1005617107
- Hellems, J., Mortier, G., De Paepe, A., Speleman, F., Vandesompele, J., 2007. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol.* 8, R19. doi:10.1186/gb-2007-8-2-r19
- Holloway, A., Strand, M., Black VI, W., Antolin, M., 2000. Linkage Analysis of Sex Determination in *Bracon sp.* Near *hebetor* (Hymenoptera: Braconidae). *Genetics* 154, 205–212.
- Huber, J., 2009. Biodiversity of Hymenoptera, in: Footitt, R., Adler, P. (Eds.), *Insect Biodiversity: Science and Society*. Willey-Blackwell, Oxford, UK. doi:10.1002/9781444308211.ch12
- Hudson, M., 2008. Sequencing breakthroughs for genomic ecology and evolutionary biology. *Mol. Ecol. Resour.* 8, 3–17.

- Hunt, G.J., Page, R.E., 1995. Linkage map of the honey Bee, based on RAPD markers. *Genetica* 1371–1382.
- Infante, C., Matsuoka, M.P., Asensio, E., Cañavate, J.P., Reith, M., Manchado, M., 2008. Selection of housekeeping genes for gene expression studies in larvae from flatfish using real-time PCR. *BMC Mol. Biol.* 12, 1–12. doi:10.1186/1471-2199-9-28
- Jaworski, D.C., Zou, Z., Bowen, C.J., Wasala, N.B., Madden, R., Wang, Y., Kocan, K.M., Jiang, H., Dillwith, J.W., 2010. Pyrosequencing and characterization of immune response genes from the American dog tick, *Dermacentor variabilis* (L.). *Insect Mol. Biol.* 19, 617–630.
- Jenkins, C., Smart, J., Fell, S., Galea, F., Marsh, I., Reynolds, O., 2011. Molecular characterisation and development of diagnostic tools for the detection of parasitoid wasp species (Hymenoptera: Braconidae) parasitising the Queensland fruit fly, *Bactrocera tryoni* (Froggatt) (Diptera: Tephritidae).
- Julsirikul, D., Worapong, J., Kitthawee, S., 2014. Analysis of mitochondrial COI sequences of the *Diachasmimorpha longicaudata* (Hymenoptera: Braconidae) species complex in Thailand. *Entomol. Sci.* 17, 231–239.
- Kitthawee, S., 2000. Seasonal Occurrence of *Diachasmimorpha longicaudata* ( Ashmead ) ( Hymenoptera : Braconidae ), a Parasitoid of *Bactrocera correcta* ( Bezzi ) ( Diptera : Tephritidae ) in a Guava Orchard in Central Thailand. *Scienceasia* 26, 87–92.
- Kitthawee, S., Dujardin, J., 2009. The *Diachasmimorpha longicaudata* complex : reproductive isolation and geometric patterns of the wing. *Biol. Control* 51, 191–197. doi:10.1016/j.biocontrol.2009.06.011
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., Marra, M. a, 2009. Circos: An Information Aesthetic for Comparative Genomics. *Genome Res.* 19, 1639–45.
- Kubista, M., Andrade, M., Lind, K., Sindelka, R., Sjo, R., Stro, L., Sta, A., 2006. The real-time polymerase chain reaction. *Mol. Aspects Med.* 27, 95–125. doi:10.1016/j.mam.2005.12.007
- Kumar, S., Shah, N., Garg, V., Bhatia, S., 2014. Large scale in-silico identification and characterization of simple sequence repeats (SSRs) from de novo assembled transcriptome of *Catharanthus roseus* (L.) G. Don. *Plant Cell Rep.* 33, 905–18. doi:10.1007/s00299-014-1569-8
- Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., Salzberg, S.L., 2004. Versatile and Open Software for Comparing Large Genomes. *Genome Biol.* 5, 12.
- Lacadena, J., 1988. *Citogenética*, 4th ed. AGESA, Madrid.
- Lanzavecchia, S., 2007. II Reunión Argentina de Parasitoidólogos “RAP 2007”.
- Lanzavecchia, S., Carabajal Paladino, L., Zandomeni, R., Cladera, J., 2008. A genomic region in *Diachasmimorpha longicaudata* (Hymenoptera: Braconidae) characterized based on sequences involved in sex determination in honeybee.

- LaSalle, J., Gauld, I., 1993. Hymenoptera and biodiversity. CAB International.
- Lasalle, J., Gauld, I., 1991. Parasitic Hymenoptera and the biodiversity crisis. *Redia* 74, 315–334.
- Lawrence, P., 2005a. Morphogenesis and cytopathic effects of the *Diachasmimorpha longicaudata* entomopoxvirus in host haemocytes. *J. Insect Physiol.* 51, 221–233. doi:doi:10.1016/j.jinsphys.2004.12.003
- Lawrence, P., 2005b. Transmission of the *Diachasmimorpha longicaudata* rhabdovirus (DIRhV) to wasp offspring: an ultrastructural analysis. *J. Insect Physiol.* 51, 235–24. doi:doi:10.1016/j.jinsphys.2005.01.002
- Lawrence, P., 1982. *Biosteres longicaudatus*: development dependence on host (*Anastrepha suspensa*). *Physiol. Exp. Parasitol.* 53, 396–405.
- Lawrence, P., 1981. Host vibration: a cue to host location by the parasite *Biosteres longicaudatus*. *Oecologia* 249–251.
- Lawrence, P., Baranowsky, R., Greany, P., 1976. Effect of host age on development of *Biosteres* (*Opius*) *longicaudatus*, a parasitoid of the Caribbean fruit fly *Anastrepha suspensa*. *Florida Entomol.* 59, 33–39.
- Leyva, V., Martínez, L., Valdez, J., 1988. Mecanismos de detección del huésped utilizados por *Diachasmimorpha* (*Biosteres*) *longicaudatus* (Ashmead) (Hymenoptera: Braconidae) y estructura de los órganos de recepción. *Folia Entomológica Mex.* 76, 165–176.
- Liendo, M., Devescovi, F., Bachmann, G., Utgés, M., Abraham, S., Vera, M., Lanzavecchia, S., Bouvet, J., Gómez-Cendra, P., Hendrichs, J., Teal, P., Cladera, J., Segura, D., 2012. Precocious sexual signalling and mating in *Anastrepha fraterculus* (Diptera: Tephritidae) sterile males achieved through juvenile hormone treatment and protein Supplements. *Bull. Entomol. Res. Working Paper*.
- López, M., Aluja, M., Sivinski, J., 1999. Hymenopterous larval–pupal and pupal parasitoids of *Anastrepha* flies (Diptera: Tephritidae) in Mexico. *Biol. Control* 15, 119–129.
- Lourenço, A., Mackert, A., Cristino, A., Simões, Z., 2008. Validation of reference genes for gene expression studies in the honey bee, *Apis mellifera*, by quantitative real-time RT-PCR. *Apidologie* 39, 372–385. doi:10.1051/apido:2008015
- Manoukis, N., Geib, S., Seo, D., McKenney, M., Vargas, R., Jang, E., 2011. An optimized protocol for rearing *Fopius arisanus*, a parasitoid of tephritid fruit flies. *JoVE J.* 53, e2901. doi:10.3791/2901
- Manu, Ludwig, M., Kreitman, M., 2013. Sex-specific Pattern Formation During Early *Drosophila* Development. *Genetics* 194, 1–37. doi:10.1534/genetics.112.148205
- Marin, I., Siegal, M., Baker, B., 2000. The evolution of dosage-compensation mechanisms. *BioEssays* 22, 1106–1114.

- Martínez, J., 2009. Diversidad de Doryctinae (Insecta: Hymenoptera: Braconidae) de la Argentina, con énfasis en las formas galícolas. Universidad de Buenos Aires.
- Martínez, L., 1992. Parasitismo, supervivencia y reproducción de *Diachasmimorpha longicaudata* parasitoide de moscas de la fruta. Colegio de Postgraduados, Montecillos, México.
- Martínez, L., 1989. Factores de importancia para la cría masiva de *Diachasmimorpha longicaudata* (Ashmead) (Hymenoptera: Braconidae)., in: XXV Congreso Nacional de Entomología. Oaxaca, México.
- Marucci, P., Clancy, D., 1950. The artificial culture of fruit flies and their parasites. Proc. Hawaiian Entomol. Soc. 14, 163–166.
- Mita, K., Morimyo, M., Okano, K., Koike, Y., Nohata, J., Kawasaki, H., Kadono-Okuda, K Yamamoto, K., Suzuki, M., Shimada, T., Goldsmith, M., Maeda, S., 2003. The construction of an EST database for *Bombyx mori* and its application. Proc. Natl. Acad. Sci. 100, 14121–14126.
- Mittapalli, O., Bai, X., Mamidala, P., Rajarapu, S., Bonello, P., Herms, D., 2010. Tissue-specific transcriptomics of the exotic invasive insect pest emerald ash borer (*Agrilus planipennis*). PLoS One 5, e13708. doi:10.1371/journal.pone.0013708
- Montoya, P., Cancino, J., Ruiz, L., 2012. Packing of Fruit Fly Parasitoids for Augmentative Releases. Insects 3, 889–899. doi:10.3390/insects3030889
- Montoya, P., Cancino, J., Zenil, M., Gutiérrez, J., Santiago, G., 2007. The augmentative biological control component in the Mexican campaign against *Anastrepha* fruit flies., in: Vreysen MJB, Robinson AS, H.J. (eds) (Ed.), Area-Wide Control of Insects Pests: From Research to Field Implementation. Springer, Dordrecht, pp. 661–670.
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., Kanehisa, M., 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35, W182–5. doi:10.1093/nar/gkm321
- Nishimura, O., Brillada, C., Yazawa, S., Maffei, M.E., Arimura, G., 2012. Transcriptome pyrosequencing of the parasitoid wasp *Cotesia vestalis*: genes involved in the antennal odorant-sensory system. PLoS One 7, e50664. doi:10.1371/journal.pone.0050664
- Normark, B.B., 2003. The Evolution of Alternative Genetic Systems in Insects. Annu. Rev. Entomol. 48, 397–423. doi:10.1146/annurev.ento.48.091801.112703
- O’Neil, S.T., Dzurisin, J.D.K., Carmichael, R.D., Lobo, N.F., Emrich, S.J., Hellmann, J.J., 2010. Population-level transcriptome sequencing of nonmodel organisms *Erynnis propertius* and *Papilio zelicaon*. BMC Genomics 11, 310. doi:10.1186/1471-2164-11-310
- Olafson, P., Lhomeyer, K., Dowd, S., 2010. Analysis of expressed sequence tags from a significant livestock pest, the stable fly (*Stomoxys calcitrans*), identifies transcripts with a putative role in chemosensation and sex determination. Arch. Insect Biochem. Physiol. 74, 179–204.

- Oliveira, D., Werren, J., Verhulst, E., Giebel, J., Kamping, A., Beukeboom, L., van de Zande, L., 2009. Identification and characterization of the doublesex gene of *Nasonia*. *Insect Mol. Biol.* 18, 315–324. doi:10.1111/j.1365-2583.2009.00874.x
- Oroño, L., Ovruski, S., 2007. Presence of *Diachasmimorpha longicaudata* (Hymenoptera: Braconidae) in a guild of parasitoids attacking *Anastrepha fraterculus* (Diptera: Tephritidae) in northwestern Argentina. *Florida Entomol.* 90, 2000–2002.
- Ovruski, S., Colin, C., Soria, A., Oroño, L., Schliserman, P., 2003a. Introducción y producción en laboratorio de *Diachasmimorpha tryoni* y *Diachasmimorpha longicaudata* (Hymenoptera: Braconidae) para el control biológico de *Ceratitis capitata* (Diptera: Tephritidae). *Rev. Soc. Entomológica Argentina* 62, 49–59.
- Ovruski, S., Colin, C., Soria, A., Oroño, L.E., Schliserman, P., Collin, C., Sergio, M., 2003b. Introducción y producción en laboratorio de *Diachasmimorpha tryoni* y *Diachasmimorpha longicaudata* (Hymenoptera : Braconidae ) para el control biológico de *Ceratitis capitata* (Diptera : Tephritidae ) en la Argentina. *Rev. Soc. Entomológica Argentina* 62, 49–59.
- Ovruski, S., Fidalgo, P., 1994. Use of parasitoids (Hym.) in the control of fruit flies (Dip.: Tephritidae) in Argentina: bibliographic review (1937-1991). *IOBC/WPRS Bull.* 17, 84–92.
- Ovruski, S.M., Schliserman, P., 2012. Biological Control of Tephritid Fruit Flies in Argentina: Historical Review, Current Status, and Future Trends for Developing a Parasitoid Mass-Release Program. *Insects* 3, 870–888. doi:10.3390/insects3030870
- Papanicolaou, A., Joron, M., McMillan, W., Blaxter, M., Jiggins, C., 2005. Genomic tools and cDNA derived markers for butterflies. *Mol. Ecol.* 14, 2883–2897.
- Papanicolaou, A., Stierli, R., French-Constant, R.H., Heckel, D.G., 2009. Next generation transcriptomes for next generation genomes using est2assembly. *BMC Bioinformatics* 10, 447. doi:10.1186/1471-2105-10-447
- Paranhos, B.J., Costa, M.D.L.Z., Ovruski, S.M., Alves, R.M., Blummer, L., Walder, J.M.M., 2008. Offspring in Response to Parental Female Densities in the Fruit Fly Parasitoid *DIACHASMIMORPHA LONGICAUDATA* (Hymenoptera : Braconidae : Opiinae ). *Florida Entomol.* 91, 628–635.
- Parra, G., Bradnam, K., Korf, I., 2007. Genome analysis CEGMA : a pipeline to accurately annotate core genes in eukaryotic genomes 23, 1061–1067. doi:10.1093/bioinformatics/btm071
- Paschall, J., Oleksiak, M., VanWye, J., Roach, J., Whitehead, J., Wyckoff, G., Koll, K., Crawford, D., 2004. FunnyBase: a systems level functional annotation of *Fundulus* ESTs for the analysis of gene expression. *BMC Genomics* 5, 96.
- Pauchet, Y., Wilkinson, P., van Munster, M., Augustin, S., Pauron, D., French-Constant, R.H., 2009. Pyrosequencing of the midgut transcriptome of the poplar leaf beetle *Chrysomela tremulae* reveals new gene families in Coleoptera. *Insect. Insect Biochem. Mol. Biol.* 39, 403–413.

- Pauchet, Y., Wilkinson, P., Vogel, H., Nelson, D.R., Reynolds, S.E., Heckel, D.G., ffrench-Constant, R.H., 2010. Pyrosequencing the *Manduca sexta* larval midgut transcriptome: messages for digestion, detoxification and defence. *Insect Mol. Biol.* 19, 61–75. doi:10.1111/j.1365-2583.2009.00936.x
- Pearson, D., Cassola, F., 1992. World-wide species richness patterns of tiger beetles (Coleoptera: Cicindelidae): indicator taxon for biodiversity and conservation studies. *Conserv. Biol.* 6, 376–391.
- Pemberton, CE; Willard, W., 1918. A contribution to the biology of fruit fly parasites in Hawaii. *J. Agric. Res.* 15, 419–436.
- Purcell, M., 1998. Contribution of biological control to integrated pest management of tephritid fruit flies in the tropic and subtropics. *Integr. Pest Manag. Rev.* 3, 63–83.
- Ridgeway, J. a., Timm, a. E., 2014. Comparison of RNA Isolation Methods From Insect Larvae. *J. Insect Sci.* 14, 268–268. doi:10.1093/jisesa/ieu130
- Rong, S., Li, D.-Q., Zhang, X.-Y., Li, S., Zhu, K.Y., Guo, Y.-P., Ma, E.-B., Zhang, J.-Z., 2013. RNA interference to reveal roles of  $\beta$ -N -acetylglucosaminidase gene during molting process in *Locusta migratoria*. *Insect Sci.* 20, 109–119. doi:10.1111/j.1744-7917.2012.01573.x
- Roulin, A.C., Wu, M., Pichon, S., Arbore, R., Kühn-Bühlmann, S., Kölliker, M., Walser, J.-C., 2014. De novo transcriptome hybrid assembly and validation in the European earwig (Dermaptera, Forficula auricularia). *PLoS One* 9, e94098. doi:10.1371/journal.pone.0094098
- Rudd, S., 2003. Expressed sequence tags: alternative or complement to whole genome sequences? *Trends Plant Sci.* 8, 321–329.
- Sánchez, L., 2008. Sex-determining mechanisms in insects. *Int. J. Dev. Biol.* 52, 837–56. doi:10.1387/ijdb.072396ls
- Scannapieco, A., Mannino, M., Parreño, M., Martinez, A., Lanzavecchia, S., Palacio, M., Cladera, J., 2011. Genetic basis of hygienic behavior in *Apis mellifera* (Hymenoptera: Apidae): a candidate gene approach, XXIV International Congress of Entomology. Daegu, Korea.
- Schliserman, P., Ovruski, S., De Coll, O., 2003a. The stablishment of *Diachasmimorpha longicaudata* (HYymenoptera: Braconidae) in Misiones, Northeastern Argentina. *Florida Entomol.* 86, 491–492.
- Schliserman, P., Ovruski, S., DeColl, O., 2003b. The recovery and permanent establishment of *Diachas- mimorpha longicaudata* (Hymenoptera: Braconidae) in Misiones, northeastern Argentina. *Florida Entomol.* 86, 491–492.
- Schwartz, T.S., Tae, H., Yang, Y., Mockaitis, K., Van Hemert, J.L., Proulx, S.R., Choi, J.-H., Bronikowski, A.M., 2010. A garter snake transcriptome: pyrosequencing, de novo assembly, and sex-specific differences. *BMC Genomics* 11, 694. doi:10.1186/1471-2164-11-694

- Segura, D., Vera, M., Cagnotti, C., Vaccaro, N., De Coll, O., Ovruski, S., Cladera, J., 2006. Relative Abundance of *Ceratitis capitata* and *Anastrepha fraterculus* (Diptera: Tephritidae) in Diverse Host Species and Localities of Argentina. *Ann. Entomol. Soc. Am.* 99, 70–83.
- Segura, D., Viscarret, M., Ovruski, S., Cladera, J., 2012. Response of the fruit fly parasitoid *Diachasmimorpha longicaudata* to host and host-habitat volatile cues. *Entomol. Exp. Appl.* 143, 164–176. doi:10.1111/j.1570-7458.2012.01246.x
- Shaw, S., 1995. Braconidae, in: Hanson, P., Gauld, I. (Eds.), *The Hymenoptera of Costa Rica*. Oxford University Press, Oxford, UK, pp. 431–463.
- Sindelka, R., Ferjentsik, Z., Jonák, J., 2006. Developmental expression profiles of *Xenopus laevis* reference genes. *Dev. Dyn.* 235, 754–8. doi:10.1002/dvdy.20665
- Sivinski, J., 1996. The past and potential of biological control of fruit flies. En: *Fruit fly pests: A world assessment of their biology and management.*, in: McPherson BA, S.G. (Ed.), . St. Lucie. Del Ray Beach, Florida, pp. 365–375.
- Sivinski, J., Pinero, J., Aluja, M., 2000. The distribution of parasitoids (Hymenoptera) of *Anastrepha* fruit flies (Diptera: Tephritidae) along an altitudinal gradient in Veracruz, Mexico. *Biol. Control* 18, 258–269.
- Smith, C.R., Smith, C.D., Robertson, H.M., Helmkampf, M., Zimin, A., Yandell, M., Holt, C., Hu, H., Abouheif, E., Benton, R., Cash, E., Croset, V., Currie, C.R., Elhaik, E., Elsik, C.G., Favé, M.-J., Fernandes, V., Gibson, J.D., Graur, D., Gronenberg, W., Grubbs, K.J., Hagen, D.E., Viniegra, A.S.I., Johnson, B.R., Johnson, R.M., Khila, A., Kim, J.W., Mathis, K. a, Munoz-Torres, M.C., Murphy, M.C., Mustard, J. a, Nakamura, R., Niehuis, O., Nigam, S., Overson, R.P., Placek, J.E., Rajakumar, R., Reese, J.T., Suen, G., Tao, S., Torres, C.W., Tsutsui, N.D., Viljakainen, L., Wolschin, F., Gadau, J., 2011. Draft genome of the red harvester ant *Pogonomyrmex barbatus*. *Proc. Natl. Acad. Sci. U. S. A.* 108, 5667–72. doi:10.1073/pnas.1007901108
- Stolle, E., Kidner, J.H., Moritz, R.F. a, 2013. Patterns of evolutionary conservation of microsatellites (SSRs) suggest a faster rate of genome evolution in Hymenoptera than in Diptera. *Genome Biol. Evol.* 5, 151–62. doi:10.1093/gbe/evs133
- Stuhl, C., Cicero, L., Sivinski, J., Teal, P., Lapointe, S., Paranhos, B., Aluja, M., 2011. Longevity of multiple species of tephritid (Diptera) fruit fly parasitoids (Hymenoptera: Braconidae: Opiinae) provided exotic and sympatric-fruit based diets. *J. Insect Physiol.* 57, 1463–1470.
- Suarez, L., Murua, F., Lara, N., Escobar, J., Taret, G., Rubio, J.L., Nieuwenhove, G. Van, Bezdjian, L., Schliserman, P., Ovruski, S.M., 2014. Biological Control of *Ceratitis capitata* (Diptera : Tephritidae) in Argentina : Releases of *Diachasmimorpha longicaudata* (Hymenoptera : Braconidae) in Fruit-Producing Semi-Arid Areas of San Juan. *Nat. Sci.* 6, 664–675.
- Tejangkura, T., Kitthawee, S., Chungjatupornchai, W., 2004. Mariner transposable elements of *Diachasmimorpha longicaudata* strain A.
- Terán, H., 1977. Comportamiento alimentario y su correlación a la reproducción en hembras de *Ceratitis capitata* (Wied.) (Diptera, Trypetidae). *Rev. Agronómica N.O.A.* 14, 17–34.

- Terán López, E., 1983. Efecto de la edad del huésped sobre la preferencia y el parasitismo de *Biosteres longicaudatus* en larvas de *Anastrepha ludens* (Loew) y *Musca domestica* (Linn.). Universidad de Tamaulipas, México.
- Turica, A., 1968. Lucha biológica como medio de control de las moscas de los frutos. IDIA 241, 29–38.
- Van de Zande, L., Verhulst, E.C., 2014. Genomic imprinting and maternal effect genes in haplodiploid sex determination. *Sex Dev.* 8, 74–82. doi:10.1159/000357146
- Vargas, R., Leblanc, L., Harris, E., Manoukis, N., 2012. Regional Suppression of *Bactrocera* Fruit Flies (Diptera: Tephritidae) in the Pacific Through Biological Control and Prospects for Future Introductions into Other Areas of the World. *Insects* 3, 727–742.
- Vargas, R., Ramadan, M., Hussain, T., Mochizuki, N., Bautista, R., Stark, J., 2002. Comparative demography of six fruit (Diptera: Tephritidae) parasitoids (Hymenoptera: Braconidae). *Biol. Control* 25, 30–40.
- Vera, J., Wheat, C., Fescemyer, H., Frilader, M., Crawford, D., 2008. Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Mol. Ecol.* 17, 1636–1647. doi:10.1111/j.1365-294X.2008.03666.x
- Verhulst, E.C., Beukeboom, L.W., van de Zande, L., 2010. Maternal control of haplodiploid sex determination in the wasp *Nasonia*. *Science* 328, 620–3. doi:10.1126/science.1185805
- Verhulst, E.C., Lynch, J. a, Bopp, D., Beukeboom, L.W., van de Zande, L., 2013. A new component of the *Nasonia* sex determining cascade is maternally silenced and regulates transformer expression. *PLoS One* 8, e63618. doi:10.1371/journal.pone.0063618
- Viscarret, M., Carabajal Paladino, L., Jur, i M., Cladera, J., 2006a. Dosis de irradiación para la obtención de mutantes en el parasitoide *Diachasmimorpha longicaudata* (Ashmead) (Hymenoptera: Braconidae), in: XXXV Congreso Argentino de Genética SAG. San Luis, Argentina.
- Viscarret, M., La Rossa, R., Segura, D., Ovruski, D., Cladera, J., Ovruski, S., Cladera, J., 2006b. Evaluation of the parasitoid *Diachasmimorpha longicaudata* (Ashmead) (Hymenoptera : Braconidae) reared on a genetic sexing strain of *Ceratitis capitata* (Wied.) (Diptera : Tephritidae). *Biol. Control* 36, 147–153. doi:10.1016/j.biocontrol.2005.08.009
- Vitková, M., 2007. Structural and molecular differentiation of sex chromosomes in Lepidoptera.
- Weber, A., Weber, K., Carr, K., Wilkerson, C., Ohlrogge, J., 2007. Sampling the Arabidopsis transcriptome with massively parallel pyrosequencing. *Plant Physiol.* 144, 32–42.
- Werren, J., Richards, S., Desjardins, C., Niehuis, O., Gadau, J., Colbourne, J., The *Nasonia* Genome Working Group, 2009. Functional and Evolutionary Insights from the Genomes of Three Parasitoid *Nasonia* Species. *Science* (80- ). 327, 343–348.
- Whitfield, C., Band, M., Bonaldo, M., Kumar, C., Liu, L., Pardinas, J., Robertson, H., Soares, M., Robinson, G., 2002. Annotated expressed sequence tags and cDNA microarrays for



- studies of brain and behavior in the honey bee. *Genome ...* 12, 555–566.  
doi:10.1101/gr.5302.information
- Whitfield, J., 1998. Phylogeny and evolution of host-parasitoid interactions in Hymenoptera. *Annu. Rev. Entomol.* 43, 129–151. doi:10.1146/annurev.ento.43.1.129
- Wilson, E., 1987. The little things that run the world. *Conserv. Biol.* 1, 334–346.  
doi:10.1111/j.1523-1739.1987.tb00055.x
- Winnebeck, E.C., Millar, C.D., Warman, G.R., 2010. Why does insect RNA look degraded? *J. Insect Sci.* 10, 159. doi:10.1673/031.010.14119
- Yúfera, M., Halm, S., Beltran, S., Fusté, B., Planas, J. V., Martínez-Rodríguez, G., 2011. Transcriptomic Characterization of the Larval Stage in Gilthead Seabream (*Sparus aurata*) by 454 Pyrosequencing. *Mar. Biotechnol.* (NY). doi:10.1007/s10126-011-9422-3
- Zagrobelny, M., Scheibye-Alsing, K., Jensen, N.B., Møller, B.L., Gorodkin, J., Bak, S., 2009. 454 pyrosequencing based transcriptome analysis of *Zygaena filipendulae* with focus on genes involved in biosynthesis of cyanogenic glucosides. *BMC Genomics* 10, 574.  
doi:10.1186/1471-2164-10-574
- Zdobnov, E.M., Apweiler, R., 2001. InterProScan - an Integration Platform for the Signature-Recognition Methods in InterPro. *Bioinformatics* 17, 847–848.
- Zhang, F., Guo, H., Zheng, H., Zhou, T., Zhou, Y., Wang, S., Fang, R., Qian, W., Chen, X., 2010. Massively parallel pyrosequencing-based transcriptome analyses of small brown planthopper (*Laodelphax striatellus*), a vector insect transmitting rice stripe virus (RSV). *BMC Genomics* 11, 303. doi:10.1186/1471-2164-11-303
- Zhang, Y., Zheng, Y., Li, D., Fan, Y., 2014. Transcriptomics and identification of the chemoreceptor superfamily of the pupal parasitoid of the oriental fruit fly, *Spalangia endius* Walker (Hymenoptera: Pteromalidae). *PLoS One* 9, e87800.  
doi:10.1371/journal.pone.0087800
- Zou, Z., Najar, F., Wang, Y., Roe, B., Jiang, H., 2008. Pyrosequence analysis of expressed sequence tags for *Manduca sexta* hemolymph proteins involved in immune responses. *Insect Biochem. Mol. Biol.* 38, 677–682.

# ANEXOS

Alineamiento para las secuencias obtenidas a partir de ADNc para *D. longicaudata*.

Seq3,4,5 y 6: secuencias de ADNc

SeqDN: secuencia obtenida a partir de ADN genómico

Conse: secuencia de consenso utilizada para búsqueda de homología

127

[illegible]

- Secuencia variable para *csd*

SeqAD: secuencia obtenida a partir de ADN genómico

Conse: secuencia de consenso utilizada para búsqueda de homología

510 520 530 540 550 560 570 580 590 600

seq4. GTGATGCTACTC

seq5. GTGATGCTACTCGACAGANNNN

[illegible]

- Secuencias utilizadas para el diseño de cebadores inespecíficos del gen *dsx*

Alineamiento de proteínas DSX: regiones conservadas y dominios funcionales. Se muestran resaltadas en amarillo las regiones conservadas dentro de los dominios funcionales. En verde se resaltan las regiones sexo-específicas para las diferentes especies.

- gi|15: *A. mellifera*, *dsx* variante femenina
- gi|15: *A. mellifera*, *dsx* variante masculina
- gi|59: *Drosophila melanogaster*, *dsx* variante femenina
- gi|80: *M. rotundata*, *dsx* variante X1
- gi|24: *N. vitripennis*, *dsx* variante femenina
- gi|24: *N. vitripennis*, *dsx* variante masculina
- gi|21: *T. dubius*, *dsx* variante masculina
- pfam0: pfam00751 dominio de unión al ADN, secuencia consenso (dominio DM).
- pfam0: pfam00751 variante 1 de la secuencia del dominio de unión al ADN
- pfam0: pfam00751 variante 2 de la secuencia del dominio de unión al ADN
- pfam0: pfam08828 dominio de dimerización de *dsx*, secuencia conservada (dominio *dsx* “Dimer”).

```

10      20      30      40      50      60      70      80      90      100     110
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
gi|15MYREENEQNRAADLAPQQPSGANTFERLEHSQDSKNGDDGPKKVQTDASSSTN-TPKPR-ARNCARCLNHL-EITLKSHKRY-CKYRTCTCEKCKITANRQQVMRQNMK
gi|15MYREENEQNRAADLAPQQPSGANTFERLEHSQDSKNGDDGPKKVQTDASSSTN-TPKPR-ARNCARCLNHL-EITLKSHKRY-CKYRTCTCEKCKITANRQQVMRQNMK
gi|59-----MVSEENWNSDTMSDSMDIDSKNDVCGGASSSSGSSISPRTPPNCARCRNHGL-KITLKGHKRY-CKFRYCTCEKCRILTADRQVRMALQTA
gi|80-----MDNGDDS-----ASAQLEKSSSSSTNGAPNPKPPNCARCRNHQL-KIAVKGHKRY-CKYRYCTCDSCKLTADRQVRMARQTA
gi|24-----MDQSDDMVS-----SDRESRLGQTSTKKPKPSQR-IPKCTRCQNHGK-KVQVKFHKRE-CEFRYCLCEMCILTTRQQIMKVQTA
gi|24-----MDQSDDMVS-----SDRESRLGQTSTKKPKPSQR-IPKCTRCQNHGK-KVQVKFHKRE-CEFRYCLCEMCILTTRQQIMKVQTA
gi|21-----MDQSDDMVT-----SDRESRLGQTSTKKPKPSQR-IPQCTRCQNHGK-KVRVKFHKRE-CEFRYCLCEMCILTTRQQIMKVQTA
pfam0-----R-SPKCARCRNHGV-LSPLKGHKRY-CPYRDCTCEKCALVAERQRLMAAQ--
pfam0-----R-IPYCQRCENHGE-KVPLKGHKRY-CPFRDCPCKKCTLVEKRRRLMALQ--
pfam0-----SRSPIDVL-MKIFPNKRKSVLELVLRKCKGDVLQAIELAL--
pfam0-----

120      130      140      150      160      170      180      190      200     210     220
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
gi|15LKRHLAQDKVKVRVA-EEVDPLPFGV-----ENTISSVPQPPRSLEGSYDSSSGDSPVSSSSNGIHT---GFGGS
gi|15LKRHLAQDKVKVRVA-EEVDPLPFGV-----ENTISSVPQPPRSLEGSYDSSSGDSPVSSSSNGIHT---GFGGS
gi|59LRRRAQADEQRALHMHEVPPANPAATTLLSHHHHVAAPAHVHAHHVHAHHAGGHHSHHGHVHLHHQAAAAAAAAAPSAPASHLGGSSTAASSIHGHAHAHHVHMAAAAAA
gi|80LRRALALDERRVRTDDEQVDPLPFGA-----EDRLPSVLQPARSLEGSYDSSSGDSPISNHGNNVHN---NLN--
gi|24QRRARQQHEMLMEMRKKSAK-----SKDSETPPAPSMNGSNDNSSDGDMRSLSNNNNSNSETCRSIGEP
gi|24QRRARQQHEMLMEMRKKSAK-----SKDSETPPAPSMNGSNDNSSDGDMRSLSNNNNSNSETCRSIGEP
gi|21QRRAREQRELLKEMRKKSAT-----SEDSETPPAPSMNGSNDNSSDGDMRSLSNNNNSNSETCRSIGEP
pfam0-----
pfam0-----
pfam0-----
pfam0-----

230      240      250      260      270      280      290      300      310     320     330
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
gi|15IITIPPTRKLPPLHPHTA--MVTHLPQTLT-----
gi|15IITIPPTRKLPPLHPHTA--MVTHLPQTLT-----
gi|59SVAQHQQSHPHSHHHHHQHNNHQP HQQPATQTALRSPPHSDHGGSGVGPATSSSGGAPSSSNAAAATSSNGSSGGGGGGGGGGSSGGGAGGGRSSGTSVITSADHHMTTV
gi|80VITVSTSRKLPSPMHPHTG--STTHLPQSQN-----

```



[illegible]

## ANEXO 2.

- Alineamiento de secuencias de ARNm para algunas de las especies seleccionadas sobre el cual se diseñaron los cebadores inespecíficos para la amplificación de *dsx* en *D. longicaudata*. Se muestra solamente la porción del alineamiento sobre la cual se diseñaron los cebadores.-Cebadores: las primeras diez líneas identifican los cebadores, con F se indican los directos y con R se indican los reversos.

- gi|15:gi|159793544|gb|EU236954.1| *A. mellifera* ARNm doublesex variante masculina
- gi|15:gi|159793550|gb|EU236957.1| *A. mellifera* ARNm doublesex variante femenina
- gi|21:gi|215398411|gb|FJ440997.1| *N. vitripennis* ARNm doublesex variante masculina
- gi|21:gi|215398405|gb|FJ440994.1| *N. vitripennis* ARNm doublesex variante femenina

```

10      20      30      40      50      60      70      80      90      100
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
dsxF2-----
dsxFC-----
dsxFB-----
dsxF1-----
dsxR2-----
dsxRV-----
dsxR3-----
dsxR1-----
dsxR1-----
dsxR6-----
gi|15CACGTTTCTAGATTCTTCTCTGCATATTACCATACTCATTTTCGATCAGTTCCAATCATTGGAGATTCA-TCAAATTTCTCCACACTGTTGTGCGGTATT
gi|15CACGTTTCTAGATTCTTCTCTGCATATTACCATACTCATTTTCGATCAGTTCCAATCATTGGAGATTCA-TCAAATTTCTCCACACTGTTGTGCGGTATT
gi|21-----AAGTAACCACCAAGACCGCTTGAATAAAAACAGCGCAACGATACGAGCCAGTTTG
gi|21-----AAGTAACCACCAAGACCGCTTGAATAAAAACAGCGCAACGATACGAGCCAGTTTG

      110      120      130      140      150      160      170      180      190      200
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

```

135

```

dsxF1-----
dsxR2-----
dsxRV-----
dsxR3-----AGTATGGATCAGTCGGATGACATGGTG
dsxR1-----
dsxR1-----
dsxR6-----
gi|15AGCTGTTGTTTCCCGTTACAT-GAGGGGGAGAAGCAGAAGTAGGAGGAGGATGGAATGACACGCGGACACGAGAAAGACGACGAGAGAAGACAATTGAAG
gi|15AGCTGTTGTTTCCCGTTACAT-GAGGGGGAGAAGCAGAAGTAGGAGGAGGATGGAATGACACGCGGACACGAGAAAGACGACGAGAGAAGACAATTGAAG
gi|21AAGTGTG--CACGGTGCCTCGTAGGTGTTTAGC--GTAGACAGTTAATGTGTTCTAGATGAAAGCGATAGTATGGATCAGTCGGATGACATGGTG
gi|21AAGTGTG--CACGGTGCCTCGTAGGTGTTTAGC--GTAGACAGTTAATGTGTTCTAGATGAAAGCGATAGTATGGATCAGTCGGATGACATGGTG

          410      420      430      440      450      460      470      480      490      500
      ....|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|
dsxF2-----NCGATNTCAGAATNACGG
dsxFC-----N.....
dsxFB-----
dsxF1-----
dsxR2-----
dsxRV--AGCGATCGAGAGT-CGNGNC--TGGG
dsxR3-----
dsxR1-----
dsxR1-----
dsxR6-----
gi|15TGC GCGTGTCGAAAAGTGC GCGGCACGCGGACGGAAGCTCGATGTGCGTTGTTATGCTTTTTGCGGAAC TATCTAGACGTGT.TAC..GTTA.T..T..T
gi|15TGC GCGTGTCGAAAAGTGC GCGGCACGCGGACGGAAGCTCGATGTGCGTTGTTATGCTTTTTGCGGAAC TATCTAGACGTGT.TAC..GTTA.T..T..T
gi|21T-CGAGCGATCGAGAGT-CGAGGC--TGGGCCAGACATCGACGAAAAAGCCCAGCCAAGTCAGCGCATACCAAAAATGCAC.....
gi|21T-CGAGCGATCGAGAGT-CGAGGC--TGGGCCAGACATCGACGAAAAAGCCCAGCCAAGTCAGCGCATACCAAAAATGCAC.....

          510      520      530      540      550      560      570      580      590      600
      ....|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|
dsxF2CA
dsxFCN.AG--ANGGTCCA
dsxFB-----
dsxF1-----
dsxR2-----ATGTGNATACTCACGNCCAANA
dsxRV-----

```

dsxR3  
dsxR1  
dsxR1  
dsxR6  
gi|15.GATTCTACGATTTAATTTTCACATTCCACATGCGTAGTGTTTTCTAGTACAGTGATACTAGATATCGTGTTTAGGAGTTGCTTGAACGACCGGTTAAAA  
gi|15.GATTCTACGATTTAATTTTCACATTCCACATGCGTAGTGTTTTCTAGTACAGTGATACTAGATATCGTGTTTAGGAGTTGCTTGAACGACCGGTTAAAA  
gi|21..AG--AAGGTCCAGGT--CAAGTTCCACAAGCGCG-----AGTGC--GAGTTCAGATACTGTCTCTGTGAGATGTGCATACTCACGACCAAGA  
gi|21..AG--AAGGTCCAGGT--CAAGTTCCACAAGCGCG-----AGTGC--GAGTTCAGATACTGTCTCTGTGAGATGTGCATACTCACGACCAAGA

610 620 630 640 650 660 670 680 690 700  
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

dsxF2  
dsxFC  
dsxFB  
dsxF1  
dsxR2  
dsxRV  
dsxR3  
dsxR1  
dsxR1  
dsxR6  
gi|15AACTATGGGAACAGCAATGGACCAGCGAAATAGTTCGTGACGTGTGAAAAATAAAAGAAGGTGACAGAAACAAGGAAAAAATATTTATAAAACAACCTTAAC  
gi|15AACTATGGGAACAGCAATGGACCAGCGAAATAGTTCGTGACGTGTGAAAAATAAAAGAAGGTGACAGAAACAAGGAAAAAATATTTATAAAACAACCTTAAC  
gi|21GGCAGCAA-ATTATGAAGGTGCAAAACCGCTCAGAGGAGGGCGAGACAGCAACACGAGATGCTGATGGAAATGAGGAAGAAAT----CGGCCAAATCCAA  
gi|21GGCAGCAA-ATTATGAAGGTGCAAAACCGCTCAGAGGAGGGCGAGACAGCAACACGAGATGCTGATGGAAATGAGGAAGAAAT----CGGCCAAATCCAA

710 720 730 740 750 760 770 780 790 800  
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

dsxF2  
dsxFC  
dsxFB  
dsxF1  
dsxR2  
dsxRV  
dsxR3  
dsxR1  
dsxR1

CGGTGACATGCCGTAGTTTGA GCAATA A  
GACAGCAATAGCAGCGACGGTGACAT

```
dsxR6
gi|15GAGTGCGCAACGTCTCATTATTTTGAAATGTTAATCGTGATGTGCATACATGTGACATTTATATGAACGAAAAATACATGATGGAATATATAAACATGTC
gi|15GAGTGCGCAACGTCTCATTATTTTGAAATGTTAATCGTGATGTGCATACATGTGACATTTATATGAACGAAAAATACATGATGGAATATATAAACATGTC
gi|21GGACTCGGAG-ACACCACCGGCTCCAAGCATGAATGGCAGTAATGACAGCAATAGCAGCGACGGTGACATGCGTAGTTTGA-GCAATA-ACAACAGCAAC
gi|21GGACTCGGAG-ACACCACCGGCTCCAAGCATGAATGGCAGTAATGACAGCAATAGCAGCGACGGTGACATGCGTAGTTTGA-GCAATA-ACAACAGCAAC

      810      820      830      840      850      860      870      880      890      900
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

dsxF2
dsxFC
dsxFB
dsxF1
dsxR2
dsxRV
dsxR3
dsxR1
dsxR1
dsxR6
gi|15GGCTTGTCTACGCACGCGCGTGATTGTTCTCGCGCGGAATGCAATATTAAATTGGTACTGGACGCATGGCAGAAATCGTATGCGTTCTGTTTATAAAAA
gi|15GGCTTGTCTACGCACGCGCGTGATTGTTCTCGCGCGGAATGCAATATTAAATTGGTACTGGACGCATGGCAGAAATCGTATGCGTTCTGTTTATAAAAA
gi|21AGCGAGAC-GTGCAGGAGCATTTGGCGAGCCACTGC-CGAGTATACCAATACCGCAAAATTTGCCACCCC-CGCTACCCCATACGACTTCGCCCCCATCA
gi|21AGCGAGAC-GTGCAGGAGCATTTGGCGAGCCACTGC-CGAGTATACCAATACCGCAAAATTTGCCACCCC-CGCTACCCCATACGACTTCGCCCCCATCA

      910      920      930      940      950      960      970      980      990     1000
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|

dsxF2
dsxFC
dsxFB
dsxF1
dsxR2
dsxRV
dsxR3
dsxR1
dsxR1
dsxR6
gi|15GCGAGTAACACGGTAACCTTGCAATAAACTGTGAAGTATAGTGGACACAGAAAAATTGTAAATTGAGATAAAATGTGAAATATAGTGTTTTTTTTACGACT
gi|15GCGAGTAACACGGTAACCTTGCAATAAACTGTGAAGTATAGTGGACACAGAAAAATTGTAAATTGAGATAAAATGTGAAATATAGTGTTTTTTTTACGACT
```

gi|21C-----GCTCTTCGAACCCGAACCAAAATCCACGACTTGTTG-----AAGAATTATTAG--GATACAGTGTTAAATTATTACAACGGTTTGGCTACC  
gi|21C-----GCTCTTCGAACCCGAACCAAAATCCACGACTTGTTG-----AAGAATTATTAG--GATACAGTGTTAAATTATTACAACGGTTTGGCTACC

1010 1020 1030 1040 1050 1060 1070 1080 1090 1100  
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

dsxF2-----  
dsxFC-----  
dsxFB-----  
dsxF1-----  
dsxR2-----  
dsxRV-----  
dsxR3-----  
dsxR1-----  
dsxR1--GNCAATCGNTAACTNTGATGTACGTA-----  
dsxR6-----

gi|15ATTGTGAATTTGTGCTTCAAATGCCGGTTATCTATGTATTGCTGTTGATTGCTGAGAAATTTGAACATACAACGAACAATTTTCATAGAAGACAATTTTA--  
gi|15ATTGTGAATTTGTGCTTCAAATGCCGGTTATCTATGTATTGCTGTTGATTGCTGAGAAATTTGAACATACAACGAACAATTTTCATAGAAGACAATTTTA--  
gi|21ATTGGCAATCGTTAACTTTGATGTACGTAATCTTAAAGGACTCGAGGGCGGATGTGGAAGTAGCCATGCGGCGTATAACTCAAGGTAACCAATCATAAGA  
gi|21ATTGGCAATCGTTAACTTTGATGTACGTAATCTTAAAGGACTCGAGGGCGGATGTGGAAGTAGCCATGCGGCGTATAACTCAAG-----

1110 1120 1130 1140 1150 1160 1170 1180 1190 1200  
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

dsxF2-----  
dsxFC-----  
dsxFB-----  
dsxF1-----  
dsxR2-----  
dsxRV-----  
dsxR3-----  
dsxR1-----  
dsxR1-----  
dsxR6-----

gi|15--TCTGGGATTTTATCTGGGATATTGGAACGAGGAATCGGGGGAAAGAAAACCTGGTGTGCGAAAATCGAATCTACGCCTCGACTACGTTT--CGAAACAC  
gi|15--TCTGGGATTTTATCTGGGATATTGGAACGAGGAATCGGGGGAAAGAAAACCTGGTGTGCGAAAATCGAATCTACGCCTCGACTACGTTT--CGAAACAC  
gi|21ATTTACTCATAAAATATAAAGTATTAGTGCTTATCGTCAAAAGTTCAAGTATTTTTTGGCATTGCAATGATTTAAAGTTAATTTTTTCTCTAGCAAAAAAC  
gi|21-----CAAAAAAC

```

1210      1220      1230      1240      1250      1260      1270      1280      1290      1300
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
dsxF2-----
dsxFC-----
dsxFB-----
dsxF1-----
dsxR2-----
dsxRV-----
dsxR3-----
dsxR1-----
dsxR1-----
dsxR6-----
gi|15GTGTT---CTCGTTTTTTACAAGCGCGCGATAAAAGGATTAGAGAGAGAGAGAGAAAGGACAACGATAGAGGGACAAACAACCGTTCAAACATTTCAATT
gi|15GTGTT---CTCGTTTTTTACAAGCGCGCGATAAAAGGATTAGAGAGAGAGAGAGAAAGGACAACGATAGAGGGACAAACAACCGTTCAAACATTTCAATT
gi|21GTCTGGCAACCAGAACTTTATTGAGAATTATTTCTGTCTAGTCAATTTTTTATATCTAAATCGTAAACAGTTAAATATATTTTCATCCTAGTATTTCAATC
gi|21GTCTGGCAACCAGAACTTTATTGAGAATTATTTCTGTCTAGTCAATTTTTTATATCTAAATCGTAAACAGTTAAATATATTTTCATCCTAGTATTTCAATC

1310      1320      1330      1340      1350      1360      1370      1380      1390      1400
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
dsxF2-----
dsxFC-----
dsxFB-----
dsxF1-----
dsxR2-----
dsxRV-----
dsxR3-----
dsxR1-----
dsxR1-----
dsxR6-----
gi|15GAGATTGTTCTTTGTAATTATGA--AAAGGCTGTGAATCGAGGTTACCTATGTATCGCGAAGAGAACGAGCAAAACAGAGCCGCGGACTTGGCTCCCCAA
gi|15GAGATTGTTCTTTGTAATTATGA--AAAGGCTGTGAATCGAGGTTACCTATGTATCGCGAAGAGAACGAGCAAAACAGAGCCGCGGACTTGGCTCCCCAA
gi|21TTGAATGATTATTACTATCAAAAGTAATGGCTGTGAATT-----CTTGTAACCTGATGAAAATTATTTTAAAGTAACGCTGCATTGAAATAAATTCATAAT
gi|21TTGAATGATTATTACTATCAAAAGTAATGGCTGTGAATT-----CTTGTAACCTGATGAAAATTATTTTAAAGTAACGCTGCATTGAAATAAATTCATAAT

1410      1420      1430      1440      1450      1460      1470      1480      1490      1500
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
dsxF2-----
```



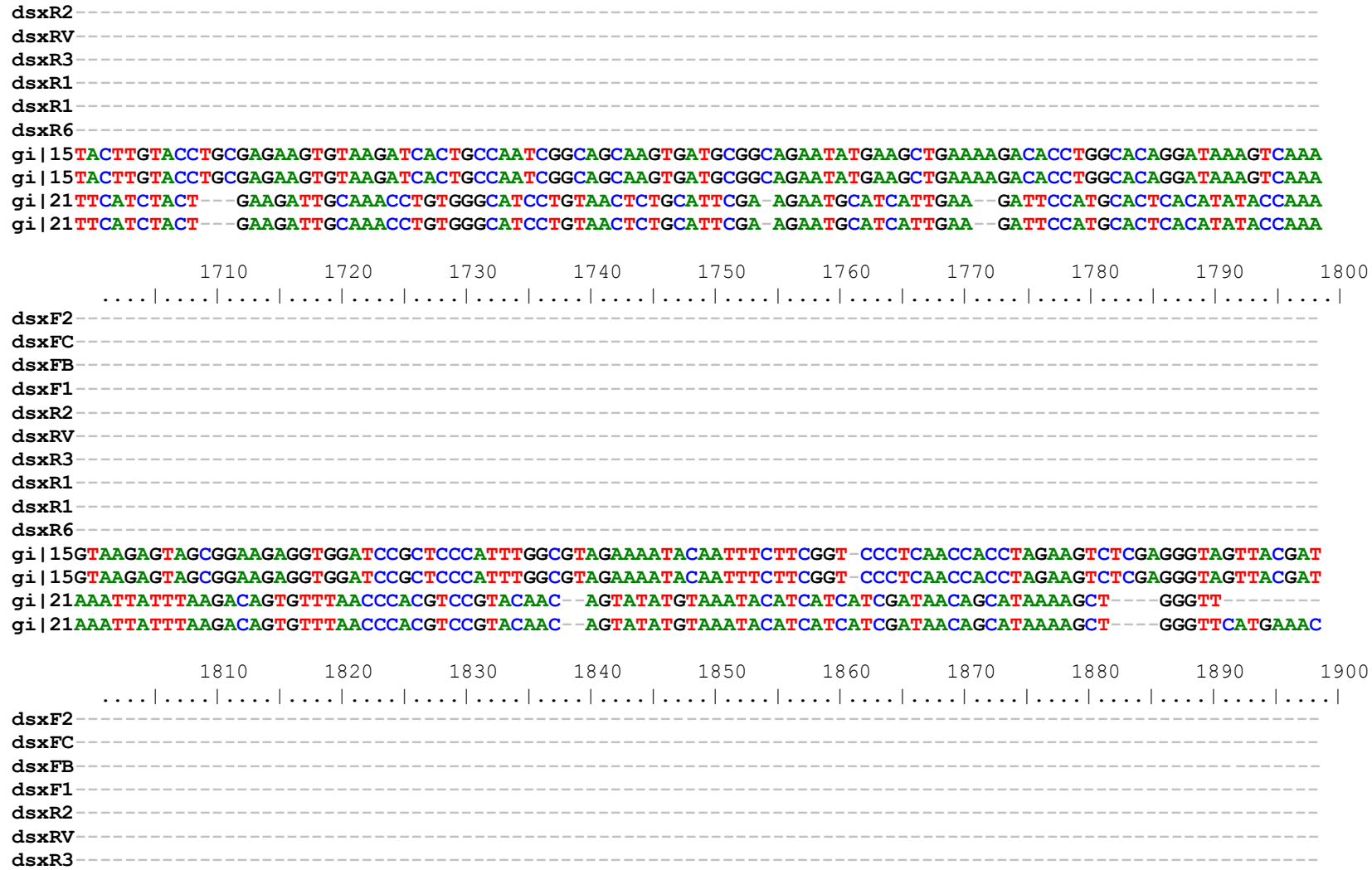
```

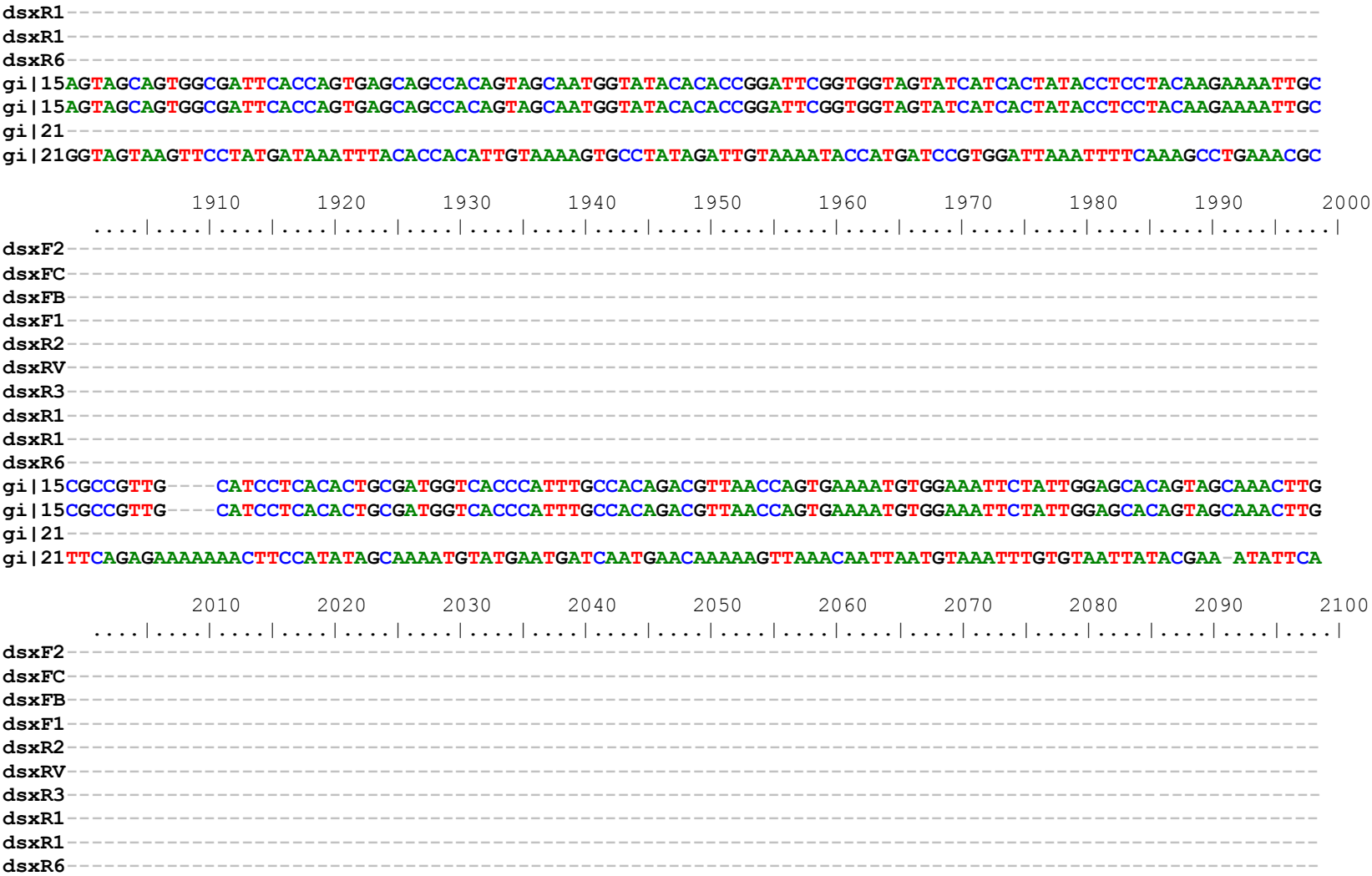
dsxFC-----
dsxFB-----
dsxF1-----
dsxR2-----
dsxRV-----
dsxR3-----
dsxR1-----
dsxR1-----
dsxR6-----
gi|15CAACCGAGTGGTGCAAAACACGTTTCGAGCGTTTGGAACTTCTCAGGATAGCAAAAAATGGGGACGATGGTCCCAGAAGGTGCAAAACAGACGCTTCCTCTT
gi|15CAACCGAGTGGTGCAAAACACGTTTCGAGCGTTTGGAACTTCTCAGGATAGCAAAAAATGGGGACGATGGTCCCAGAAGGTGCAAAACAGACGCTTCCTCTT
gi|21T--TTACTTGACACAAATGAATGC-AGTTTTTGAATAATAAGATGGTTGATGCCAATGTTCTATTACAACCTTGAAGAT-CAAAAGTTCTGCCAATCTT
gi|21T--TTACTTGACACAAATGAATGC-AGTTTTTGAATAATAAGATGGTTGATGCCAATGTTCTATTACAACCTTGAAGAT-CAAAAGTTCTGCCAATCTT

          1510      1520      1530      1540      1550      1560      1570      1580      1590      1600
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
dsxF2-----
dsxFC-----
dsxFB-----
dsxF1-----
dsxR2-----
dsxRV-----
dsxR3-----
dsxR1-----
dsxR1-----
dsxR6-----
gi|15CGACTAATACTCCAAAGCCGCGTGACGGAATTGTGCACGATGTCTGAATCATCGGCTGGAGATCACCTTAAAAATCGCACAAGAGGTACTGCAAGTACCG
gi|15CGACTAATACTCCAAAGCCGCGTGACGGAATTGTGCACGATGTCTGAATCATCGGCTGGAGATCACCTTAAAAATCGCACAAGAGGTACTGCAAGTACCG
gi|21CTCCGAAGAT--AAGACTTTTGAA-AAATTGAAATTATCATATATATTTTACGTTTGAAT-ACCTTAC--TCGTTGACATGAAGATTGTTTTTGA
gi|21CTCCGAAGAT--AAGACTTTTGAA-AAATTGAAATTATCATATATATTTTACGTTTGAAT-ACCTTAC--TCGTTGACATGAAGATTGTTTTTGA

          1610      1620      1630      1640      1650      1660      1670      1680      1690      1700
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
dsxF2-----
dsxFC-----
dsxFB-----
dsxF1-----

```





```

gi|15TAGAACCTTTTCCAATATCCTTGGGAAGCACTGTTATTGATGTACATCAATTTAAAAATATGCAGGGGCTAATCCGGAAGAAGTAGTGAGACGTATGGTTGA
gi|15TAGAACCTTTTCCAATATCCTTGGGAAGCACTGTTATTGATGTACATCAATTTAAAAATATGCAGGGGCTAATCCGGAAGAAGTAGTGAGACGTATGGTTGA
gi|21-----
gi|21TATTTGATAGGGGGTAACGTAGAACGAATATTAAATGTTTTATTTTATTCAA-----

```

### ANEXO 3.

- Alineamientos de las secuencias obtenidas para *csd* y sus mejores hits en la base de datos del transcriptoma de *D. longicaudata*.
- Secuencias alineadas con el mejor “hit” de BLASTn para la región variable de *csd* en *D. longicaudata*.

Isoti: Isotig02035, transcriptoma de *D. longicaudata*

*csd* V: Secuencia var obtenida a partir de ADN

*csd* V: Secuencia var obtenida a partir de ADNc

```

10      20      30      40      50      60      70      80      90      100
...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|
Isoti   ~~~~~GTCGTTTAGTTGGGTCTTTCACCGTAGAAAAATCCCACAATGAATAAAATTGCTAATACTT
cds V   ~~~~~Vctctaaaaataattncaaaagtcgcattaattaattatcttttaattgccagtttatgattaatcgttgaaatatttttcgattgctatttagaatact
csd V   ~~~~~
110     120     130     140     150     160     170     180     190     200
...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|
Isoti   ~~~~~TTGGCGATATTATCATTATCCTTCAGCATCGCATTGCTGCTAAGCGGCGGAATATTCCGAACGATTTGTCGTGGCTTTTTTCGCTAGAAGAATCTGATC
cds V   ~~~~~Vtttggcgataattatcattatcctcagcatcgcattgcatgctaagcggcggaatattccgaacgatttgcgtggcgttttttcgctagaagaatctgattc
csd V   ~~~~~TGGCTTTTTTCGCTAGAAGAATCTGATC
210     220     230     240     250     260     270     280     290     300
...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|
Isoti   ~~~~~AATCCAAATGAGAAATGACGATGATGATAATGAAGTTGTACAACATAAGAAAGTAAAGGCTGTGTAACGGAGGAGTGTCGTATAATAGCTGAGGCATTCT

```

```

cds Vaatccaatgaagaaaatgacgatgatgataatgaagtgtgtacaacataagaaagtaaaggctctgtgaaacggaggagtgctgtataatagctgaggcatt
csd VAATCCNATGAAGAAAATGNCNATGATNATAATGAAGTTGTACAACATAAGAAAGTAAAGGCTGTGTAACGGAGGAGTGTCTGATAATAGCTGAGGCATT

      310      320      330      340      350      360      370      380      390      400
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
IsotiCATGAGACTATGGATAAATCCGTTGATCCATGTGACGACTTTTACCAATTCGCTTGGTGGTGGGAGGAAAAGCAGCCTATTCCGAAATATGAAGCCT
cds Vccatgagactatggataaaatccgttggatccatgtgacgaacttttaccattcgttgcgggtgggtgggagggaaaagcagcctattccgaaatatgaagcc
csd VCCATGAGACTATGGATAAATCCGTTGATCCATGTGACGACTTTTACCAATTCGCTTGGTGGTGGGAGGAAAAGCAGCCTATTCCGAAATATGAAGCC

      410      420      430      440      450      460      470      480      490      500
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
IsotiCATGGAGCAGATTTCTGCAATTCCAATCGACAGTTGAGAGCCGCATTCGAGACCTATTGGAGGAGAAGTCAAGCGATGAAGAGATTCTTCCCGTCCGGCA
cds Vtcatggagcagattttctgcaattccaatcgacagttgagagccgcattcaggtaatgagtttctgtgaaattattgagattaattaacacgaagagtga
csd VTCAATGGAGCAGATTTCTGCAATTCCAATCGACAGTTGAGAGCCGCATTCGAGGTAATGAGTTTCGTTGAATTATTGAGATTAAATTAACACGAAGAGTGAG

      510      520      530      540      550      560      570      580      590      600
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
IsotiAGCAAAGAAGTGGTACCGTTCCGTGATGAATACGGAGGCTCGTAAAGCCCGGAGTCTCAAGCCTCTCGAGTCATATATTATGCAGAAATGGTGGATGGCCA
cds Vtgaatgctactcgacagaagtgaatgaaacgttaatttcatgcaattaatcatcttttcccttcgtttgtacaaatttcattggcgggttgggtttgcttgcg
csd VTGATGCTACTC

      610      620      630      640      650      660      670      680      690      700
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
IsotiATGATTATGGACGCCGAAGAGTGGCATGACGAGGACCTGTCTGGCAAGAAGTAGACGCTTTTTATACACATTTGACAGGAGAATCGACATTTTGGAAAA
cds Vaatgcctgtttacgcgcccagaaaaatttcgtaatagataaaattcgtagaagaaaagaatacngtaatttttcccttgaa
csd V

      710      720      730      740      750      760      770      780      790      800
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
IsotiTAGAACCACTATCCATGCCCGATGAGGAACCCGACAAAGGAATAATATTCCTTCCGGCATCATTGCCATTCACTGGGTCAAATCCCAACAAAAATACAGAAA
cds V
csd V

      810      820      830      840      850      860      870      880      890      900
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
IsotiCTACACAGGTGAAGACTATGAAATTTATCGAATGGTCATCgAAGGTGTTGCCGATTATTTATTGATCACACCCGTGCTAATGTTTCTGAAGAACAGTTG
cds V

```

csd V

910 920 930 940 950 960 970 980 990 1000  
 .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|  
 IsotiGAGAGGATATCGATGGAGTGATCGAACTGGAAAAATGATTTACATGGCTACTGATGGAGAAACCGCTATGGACATACTGGAGGAAATGGAAGATGATC

cds V

csd V

1010 1020 1030 1040 1050 1060 1070 1080 1090 1100  
 .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|  
 IsotiTGGaTGAGACGGACCTTGATACATTCTTCGAGTGGTGGGGTAATCAcACgAAagTCGATTAAAGATCCTCAGGCAATTATTGATTGGAGAAAAGTaCTTC

cds V

csd V

1110 1120 1130 1140 1150 1160 1170 1180 1190 1200  
 .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|  
 IsotiAACATCTATTTCGATATTGCTAATGTGGAAGTTGAGGGCAGTTTTCTTATGaCTCTaGCgTCCCTGCGATATTATCGTTTGCTGCCCAAAATATTGCGATC

cds V

csd V

1210 1220 1230 1240 1250 1260 1270 1280 1290 1300  
 .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|  
 IsotiGACGGATAAACGAACAATTGTGAATTACATTCACTGGAAATTTGTGAGCCAACTCTGAAGTACACGGTTGACGAAGTAAATGAGGTGCTGTTCGCACTC

cds V

csd V

1310 1320 1330 1340 1350 1360 1370 1380 1390 1400  
 .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|  
 IsotiATGAAGGAAGAAATATGGTGTGGAAGAACGAAAGCCCCTTTGGAGGGAATGCGTGAAGGAAGTCAAGATGACCGATGCAACCGCATTATTATTATTCATCGCCA

cds V

csd V

1410 1420 1430 1440 1450 1460 1470 1480 1490 1500  
 .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|  
 IsotiCACACATCAAAGAGAAGTCCTACAACCTCTGTCTCCGGATGATGGAACATATTTCAGCAAGAATTGAAGTGCCAAATCGGGCACTCAAAATTGGCTATCTGG

cds V

csd V

```

      1510      1520      1530      1540      1550      1560      1570      1580      1590      1600
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
IsotiGGAGGAGAAAAAGGAGGCGAAGACAAAGGTCGACGAAATGCAAACTTTATTGGATTCCCAGAGTGGTACAAAAATAAAACCGCTGTATCAATCACTAC
cds V
csd V

      1610      1620      1630      1640      1650      1660      1670      1680      1690      1700
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
IsotiAAAGGGCTAAACCATTTGGCAATGACTACCTGGACAATGTCTTGAGCTACAAGAGGTACGAGAAGAAATTGGCCATCCGCTCTTTCGGTGGATACAAAGGCA
cds V
csd V

      1710      1720      1730      1740      1750      1760      1770      1780      1790      1800
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
IsotiATGAGGCATGGATGATGGACCCCTCTAACAGTAAACGCAGCGTACGCGATGGGTATTAACTAATGGTTGTTCAGCAGTTGACTTCCAACCTCCTCTTTT
cds V
csd V

      1810      1820      1830      1840      1850      1860      1870      1880      1890      1900
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
IsotiCGGGGACAAGGTACCAAACATGTGTTAATTACGGCACTGCAGGAGTAGTAATCGGACACGAGATGGGCCATGGTTTGGATTATGCTGGGCTCGTAACAAAT
cds V
csd V

      1910      1920      1930      1940      1950      1960      1970      1980      1990      2000
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
IsotiGCCTCCATCCAGGAGAAATGGCCAAATTTTATTACCAACAAGCTAAATGCTTCATCGATCAATTCCATGAATATTATGGAGGACAAACTGTGAAACCTT
cds V
csd V

      2010      2020      2030      2040      2050      2060      2070      2080      2090      2100
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
IsotiCCCCATTAGTGACGACAGTTCAGAGACGAAAGGACAAAGGACCGTTCAGGAAATGTGGCTGATACCACCGTCTTGAAGCCGTGTTGCGTGCCTACAA
cds V
csd V

```

```

                2110      2120      2130      2140      2150      2160      2170      2180      2190      2200
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
IsotiGCGCATGCAGAAAGCATAAAGAGACCGCATGAGGAATTGAAACTTCCAGGATTCGAGGAGTACACCGACGAGCAAAATGTTCTTCATATCTTTCGCTGGGCTA
cds V
csd V

                2210      2220      2230      2240      2250      2260      2270      2280      2290      2300
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
IsotiTGGTGTGAAGTAACTACTCCGCAGTATCATAAAAGAAATAGCGTAATGAACGACGTTTCACAGCCAGGGAAATACCGAGTGTTGGGCTCAATTTCTAATA
cds V
csd V

                2310      2320      2330      2340      2350      2360      2370      2380      2390      2400
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
IsotiCTGAGGAGTTTGAAAAAATTTTCAACTGTCCGAAAGGGAGTCCCATGAACCCtGAGGACAAGTGTAGTGTTTGGATGAAACCGGAcGCCGCGTCCAAGCG
cds V
csd V

                2410      2420      2430      2440      2450      2460      2470      2480      2490      2500
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
IsotiCAGACAGCCGAAACATTCAAAAACGACATCATTTGGTCTCAGCACAAAAACTGGTGAACGACATGACCCTTTtCATCGCTAGGATGGGATTTGGCTATTTCA
cds V
csd V

                2510      2520      2530      2540      2550      2560      2570      2580      2590      2600
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
IsotiATGACTGCACATGAGAATTTATGAGCTACGAGCTAAAAGTACAAAAATCTAACTGAAACGACGTTTGGATAAGTGTAATAACTTATGGcGTCAACgTCA
cds V
csd V

                2610      2620      2630      2640      2650      2660      2670      2680      2690      2700
                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
IsotiTATTCAGTAGTTTTTTCGGTAAACGCTTAGTACTTCCCCCTTCCTCGGAATACCCCCTGATTTTACATGAACCGATGCTTTTTTAACCCCTGATTACG
cds V
csd V

                2710      2720      2730      2740      2750      2760      2770      2780      2790      2800

```



```

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Isoti TATTCGTAAACCGATTTTTTGAGGAAACAATTAGTTTCCTCTGTACGTTTTTATTGATATACCGGTCTTTGCATTACCCAGGGGCATAACAGGTATCGCCC
cds V
csd V

          2810      2820      2830      2840      2850      2860      2870      2880      2890      2900
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Isoti CAACGACAACACATGTTGGTACATTTAATTGGACAGATTGGTCTTTGCCCAATCCTTCATATAAACATGGGGATCAGTACCATATCTGTATTATAATTTC
cds V
csd V

          2910      2920      2930      2940      2950      2960      2970      2980      2990      3000
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Isoti TGAGAGGAAGGGGGGAGCTACTCTACACATTTACCGTTTTTTTGTATGGAAATAAGAATTCATACGTTAAAAAGGTGATTCTGTTATCAACATAGTG
cds V
csd V

          3010      3020      3030      3040      3050      3060      3070      3080      3090      3100
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Isoti tTGCAATATTTATATTTGAACATGAGAAAATCCCTCAAGAAGGCCAACGTACACTCCCCAATTGTCTCCTTCGTACTTCTATTCTTGTAACATTTTCCA
cds V
csd V

          3110      3120      3130      3140      3150      3160      3170      3180      3190      3200
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Isoti AATAATTTATCAGTTCTCCTATGTGCAATCGTGAAATCCAGTCAAGCATTATTTGACTTTTTAGTGCCACTATCATCTGCCCAACCCAAATGCCTCCCT
cds V
csd V

          3210      3220      3230      3240      3250      3260      3270      3280      3290      3300
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Isoti GGTATCGTTCTGGGAACATCAGGCCCAATCCGGGTTTTCCGCGATCCCGTCTAGCTCAATTAAACATTCTACTTTCTACCCCCCTTTtCAAGAACCAT
cds V
csd V

          3310      3320      3330      3340      3350      3360      3370      3380      3390
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|..

```

ANEXO 4

- Secuencias alineadas con el mejor “hit” de BLASTn para la región conservada de *csd* en *D. longicaudata*.

Sec.c: Secuencia cons obtenida a partir de ADNc

Sec.s: Secuencia cons obtenida a partir de ADN

The figure displays four genomic alignment tracks for HWR9C, showing sequence differences between the sec.c (red) and HWR9C (green) sequences. The tracks are organized into four regions, each with a scale from 10 to 100 (or 200, 300, 400) and a corresponding sequence alignment. The HWR9C sequence is shown in green, and the sec.c sequence is shown in red. The alignment tracks show the following differences:

- Region 1 (10-100):** The HWR9C sequence is identical to the sec.c sequence in this region.
- Region 2 (110-200):** The HWR9C sequence is identical to the sec.c sequence in this region.
- Region 3 (210-300):** The HWR9C sequence is identical to the sec.c sequence in this region.
- Region 4 (310-400):** The HWR9C sequence is identical to the sec.c sequence in this region.

151

Tabla A1. Detalle de secuencias expresadas diferencialmente asociadas al sexo.

152

isotig01454	4,63	80,12	0,057788318	17,30453564	Peptidase activity
isotig01516	4,63	58,55	0,079077711	12,64578834	Acid phosphatase activity
isotig01998	9,26	1130,98	0,00818759	122,1360691	-
isotig02013	23,14	246,53	0,093862816	10,65384615	Hydrolase activity on ester bonds
isotig02092	4,63	83,21	0,055642351	17,97192225	Protein dimerization activity
isotig02134	13,88	144,84	0,095829881	10,4351585	Carbohydrate metabolic process
isotig02168	4,63	49,31	0,093895762	10,65010799	COPII vesicle coat
isotig02245	13,88	172,57	0,080431129	12,43299712	-
isotig02378	9,26	95,53	0,096932901	10,31641469	Nucleotide binding
isotig02397	4,63	77,04	0,06009865	16,63930886	Protein binding
isotig02512	643,25	9,25	69,54054054	0,014380101	Hydrolase activity
isotig02593	4,63	77,04	0,06009865	16,63930886	Regulation of G2/M transition of mitotic cell cycle
isotig02652	185,11	15,41	12,01232966	0,083247799	Transferase activity
isotig02713	4,63	117,1	0,039538856	25,29157667	ARN binding
isotig02724	4,63	126,35	0,036644242	27,28941685	Kinase activity
isotig02776	4,63	154,08	0,030049325	33,27861771	Transferase activity
isotig02877	4,63	55,47	0,083468542	11,98056156	Nucleosome, ADN binding
isotig02913	46,28	3,08	15,02597403	0,066551426	Cation transmembrane transporter activity
isotig02983	9,26	95,53	0,096932901	10,31641469	Peroxisome, nucleic acid binding
isotig03014	92,55	3,08	30,0487013	0,033279308	Oxidoreductase activity
isotig03021	208,24	9,25	22,51243243	0,0444199	Oxidation-reduction process
isotig03151	4,63	1134,06	0,004082676	244,937365	Hydrolase activity
isotig03162	9,26	181,82	0,050929491	19,6349892	Polyamine biosynthetic process
isotig03165	37,02	3,08	12,01948052	0,083198271	Integral to membrane
isotig03253	4,63	70,88	0,06532167	15,30885529	Translation initiation factor activity
isotig03328	46,28	3,08	15,02597403	0,066551426	-
isotig03464	4,63	117,1	0,039538856	25,29157667	Protein binding
isotig03526	4,63	890,61	0,005198684	192,3563715	Hydrolase activity
isotig03608	32,39	3,08	10,51623377	0,095091077	Rab GTPase activator factor
isotig03671	32,39	3,08	10,51623377	0,095091077	-
isotig03827	37,02	3,08	12,01948052	0,083198271	Integral to membrane
isotig03861	32,39	3,08	10,51623377	0,095091077	Transferase activity
isotig03986	46,28	3,08	15,02597403	0,066551426	Piridoxal phosphate binding
isotig04069	4,63	61,63	0,07512575	13,31101512	Nucleus

isotig04146	55,53	3,08	18,02922078	0,055465514	Transferase activity
isotig04267	143,46	12,33	11,6350365	0,085947302	-
isotig04351	41,65	3,08	13,52272727	0,07394958	Spliceosomal complex
isotig04400	50,9	3,08	16,52597403	0,060510806	-
isotig04594	32,39	3,08	10,51623377	0,095091077	Calmodulin-dependent protein kinase activity
isotig04645	32,39	3,08	10,51623377	0,095091077	-
isotig04715	87,93	6,16	14,27435065	0,070055726	Catalytic activity, thiol oxidase
isotig04852	166,6	6,16	27,04545455	0,03697479	-
isotig04873	4,63	49,31	0,093895762	10,65010799	Clathrin adaptor complex
isotig04940	37,02	3,08	12,01948052	0,083198271	Transcription, ADN-dependent
isotig05028	203,62	6,16	33,05519481	0,030252431	Serine-type endopeptidase activity
isotig05190	78,67	6,16	12,7711039	0,078301767	-
isotig05414	41,65	3,08	13,52272727	0,07394958	-
isotig05963	124,95	3,08	40,56818182	0,02464986	Odorant binding
isotig06336	4,63	55,47	0,083468542	11,98056156	ARF GTPase activator activity
isotig06425	37,02	3,08	12,01948052	0,083198271	Integral to membrane, sulfotransferase activity
isotig06701	37,02	3,08	12,01948052	0,083198271	-
isotig06746	46,28	3,08	15,02597403	0,066551426	-
isotig07165	4,63	67,8	0,068289086	14,64362851	Transport, integral to membrane
isotig07236	4,63	70,88	0,06532167	15,30885529	Receptor, phosphatidate cytidyltransferase
isotig07264	50,9	718,03	0,070888403	14,10667976	Defense response
isotig07719	41,65	3,08	13,52272727	0,07394958	G-protein coupled receptor signaling pathway
isotig08174	41,65	3,08	13,52272727	0,07394958	-
isotig08193	9,26	123,27	0,075119656	13,31209503	Structural constituent of ribosome
isotig08255	4,63	70,88	0,06532167	15,30885529	-
isotig08256	4,63	86,29	0,053656275	18,63714903	Mitochondrial electron transport, NADH to ubiquinone
isotig08257	74,04	6,16	12,01948052	0,083198271	Integral to membrane, signal transduction
isotig08345	74,04	3,08	24,03896104	0,041599136	-
isotig08350	32,39	3,08	10,51623377	0,095091077	-

isotig08359	50,9	542,38	0,093845643	10,65579568	-
isotig08433	13,88	283,51	0,048957709	20,42579251	ATP-binding; protein serine/threonine kinase activity
isotig08435	18,51	292,76	0,063225851	15,81631551	-
isotig08441	129,57	1426,82	0,090810333	11,01196265	-

## ANEXO 6.

- Alineamiento de las secuencias pertenecientes al Isogrupo00011. Se muestran con su número correspondiente los isotigs analizados.

[illegible]



```

      210      220      230      240      250      260      270      280      290      300
00184 TGGAGTTTAAATTCACAAAACTCGAGCGAAGAAGTCAGAGATCACCTGGAGATCTACTCCATGAGCTTTTGCCTCACGCAGAAGTTCTGGAAGTTCTCA
00185 .....
00187 .....
00188 .....
00189 .....
00191 .....
00193 .....
00194 .....
00195 .....
00186 .....
00190 .....
00192 .....

      310      320      330      340      350      360      370      380      390      400
00184 TTGGCAAAAAGAGTATTTGTCACAGTCTGGAAGTTATCAGATTTTATGAGCTGCACGAAATCCGCAAAAGCCTTGGAGTTCTCGATTTTGAAGTAGTAGA
00185 .....A.....A.....
00187 .....
00188 .....
00189 .....A.....A.....
00191 .....A.....A.....
00193 .....
00194 .....A.....A.....
00195 .....A.....A.....
00186 .....A.....A.....
00190 .....A.....A.....
00192 .....A.....A.....

      410      420      430      440      450      460      470      480      490      500
00184 GCTGCTTGATTTGGGGGTAAAGGAGGAGCTCCTTGAGTTTCCCGATGAGCCAGAGATACCAGGCTTCCCAGACACACCCTCCACAGCGAGGGATGTTT
00185 .....
00187 .....
00188 .....

```

```

00189.....
00191.....
00193.....
00194.....
00195.....
00186.....
00190.....
00192.....

      510      520      530      540      550      560      570      580      590      600
      |...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|
00184GTCATCCAAGCCAATGAAC TTATGAATGATTTCAATGAGTCGTAAATATCCAGACCTGCGGCATCCAGTGATTTTCAGAA TTGCGTTGTATTCAGGGATT
00185.....
00187.....
00188.....
00189.....
00191.....
00193.....
00194.....
00195.....
00186.....
00190.....
00192.....

      610      620      630      640      650      660      670      680      690      700
      |...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|
00184CTGTCAAGTGCCCTGACAA TATTTTGAAC TCCTCCGATCGGATGTACTGAAGTGTCTCCTGTGTCTCTTGGTCATGTACCATGTATT CGAGCACAACT
00185.....
00187.....
00188.....
00189.....
00191.....
00193.....
00194.....
00195.....
00186.....
00190.....

```

159

```
00188 .....
00189 .....
00191 .....
00193 .....
00194 .....
00195 .....
00186 .....
00190 .....
00192 .....

          1010      1020      1030      1040      1050      1060      1070      1080      1090      1100
          .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00184CGTAAAGCGAGTGGATCTTGTCATATGGGAGAATGGCCCGGATTTCTTGAGCATCCCCCTGGACACCTGAGGTCTTGATAACTTTGGAATTGAGGACCGG
00185.....
00187.....
00188.....
00189.....
00191.....
00193.....
00194.....
00195.....
00186.....
00190.....
00192.....

          1110      1120      1130      1140      1150      1160      1170      1180      1190      1200
          .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00184TCTTGGGTAGTGGGGGACCCCGAGGAAGTCATTGAGGAAATCGATGATATTGTAGACGTTCAAGCCCGACATATTCAGATACTTCAGGAAATCACTGACT
00185.....
00187.....
00188.....
00189.....
00191.....
00193.....
00194.....
00195.....
00186.....
```

```
00190 .....
00192 .....

          1210      1220      1230      1240      1250      1260      1270      1280      1290      1300
          .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00184 TCCTGTTCCCTTCTGGAGCTCCATCAGCAGTAGCTTGAATTCATCGGTTCTAATGTATTGAGCAGCTTCCTGGACTTCAGCGTCGTACATCAAATACTTGA
00185 .....
00187 .....
00188 .....
00189 .....
00191 .....
00193 .....
00194 .....
00195 .....
00186 .....
00190 .....
00192 .....

          1310      1320      1330      1340      1350      1360      1370      1380      1390      1400
          .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00184 CACCAATCTTGGAGATTCTTCGATGGGTAGGAGCATGAGAAAGTCTTCTAGGTCCTGATTGGACGTTACGGCGGTGACGAGATTCTGGAAAATGAAGACC
00185 .....
00187 .....
00188 .....
00189 .....
00191 .....
00193 .....
00194 .....
00195 .....
00186 .....
00190 .....
00192 .....

          1410      1420      1430      1440      1450      1460      1470      1480      1490      1500
          .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00184 CAGGACTCTGGAGAGCAAACTCTCGATAGCCTTGACATTTCGATTCCAGCGGCCTTGGCCTTCTTGATAATGTTCTGGAACCTCCTGGTTAGCCATGAGGGT
00185 .....
```

```
00187.....
00188.....
00189.....
00191.....
00193.....
00194.....
00195.....
00186.....
00190.....
00192.....

      1510      1520      1530      1540      1550      1560      1570      1580      1590      1600
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00184ATCTACCAAGCTTGGAACTTTGGTGAGCGAAGACGTTCAACAAGATTCTTGAACAACGGAGACTTGGCCATTTTTCCTTGTAAGGGCTTTGATTGA
00185.....
00187.....
00188.....
00189.....
00191.....
00193.....
00194.....
00195.....
00186.....
00190.....
00192.....

      1610      1620      1630      1640      1650      1660      1670      1680      1690      1700
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00184TCGACTGGTAGGACAGCTATGATGTCCTTAAGGAGACCGGAGACTCCGCCGAGATTCTTGCAGATCTGAGATCTGGTTTGATGTTTGGGAGGCCTATAA
00185.....
00187.....
00188.....
00189.....
00191.....
00193.....
00194.....
00195.....
```

```

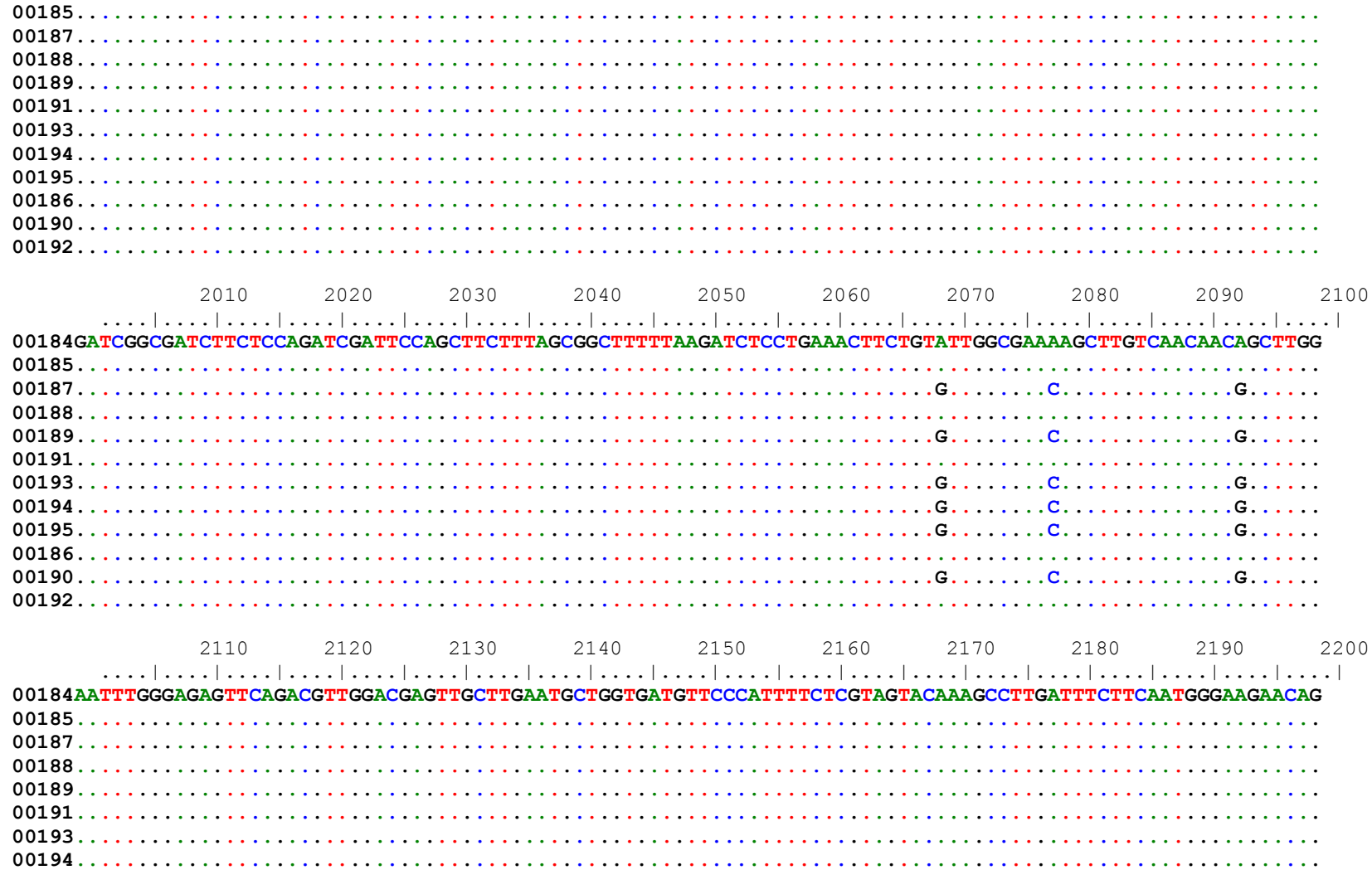
00186 .....
00190 .....
00192 .....

          1710    1720    1730    1740    1750    1760    1770    1780    1790    1800
          |.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00184 TTTTATGTAATTCGTTGATCAAGTCGTAGATTTCCAAGCCAGATTCATCTAAATAGTGAAGGAAAGCCAATACATCATCCATTTCTCAACATCACTGAC
00185 .....
00187 .....
00188 .....
00189 .....
00191 .....
00193 .....
00194 .....
00195 .....
00186 .....
00190 .....
00192 .....

          1810    1820    1830    1840    1850    1860    1870    1880    1890    1900
          |.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00184 TAGTTTCCTGAATTCCTGGGTGTGAAGATATTCCACAGCAGCTTGACCTCTTTATCTTTCCCAAATATTCTAGAAAATTTAGCGATTTTTTCTTTA
00185 .....
00187 .....
00188 .....
00189 .....
00191 .....
00193 .....
00194 .....
00195 .....
00186 .....
00190 .....
00192 .....

          1910    1920    1930    1940    1950    1960    1970    1980    1990    2000
          |.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00184 GGCAGAAAGTTTCAGGAAGTCATGGATGTCAACGTAAAGACTGCGGGCGATGCTTCTGCTATTTGGTGGGGGGAATTTTAGTCCGAATATGGTGGTTAAAA

```





```
00195 .....
00186 .....
00190 .....
00192 .....

      2210      2220      2230      2240      2250      2260      2270      2280      2290      2300
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00184 CTTTGATATCATGAATCAGACCGTTGAGTCCACCAGTGATTGTCCGTGCAACTCTGGGGTGGACTAGAGGAGGTAGACCAATGAAATCATGCAGTTTGTT
00185 .....
00187 .....T.....
00188 .....
00189 .....T.....
00191 .....
00193 .....T.....
00194 .....T.....
00195 .....T.....
00186 .....
00190 .....T.....
00192 .....

      2310      2320      2330      2340      2350      2360      2370      2380      2390      2400
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00184 GACAGCTCCGTAGATGCTAAACCCGACTCCTGGATGTACTTGTAGAAATTTTAAACGTCATCCATCTCTTCAACAACCTTGACCAGTTGCTGAACTCG
00185 .....
00187 .....
00188 .....
00189 .....
00191 .....
00193 .....
00194 .....
00195 .....
00186 .....
00190 .....
00192 .....

      2410      2420      2430      2440      2450      2460      2470      2480      2490      2500
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
```

```

00184GCACTCTGCATGTATTGACAGCTTTCTGGACGTCCCTTATCATGCTCCAGGTATTCCATCACAATATGCATAACTTCATTTTGGGGGATGAGGGCCAGAA
00185.....
00187.....
00188.....
00189.....
00191.....
00193.....
00194.....
00195.....
00186.....
00190.....
00192.....

```

```

                2510      2520      2530      2540      2550      2560      2570      2580      2590      2600
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00184AGTCTGCAAGGTCTTCCCAAGACCAAGAACCCTCACGACGCAGTTCATGTCCAGGGAATTTAATTCCAAGCAACTTGACAAGTACATCCGCGATGATATC
00185.....
00187.....
00188.....
00189.....
00191.....
00193.....
00194.....
00195.....
00186.....
00190.....
00192.....

```

```

                2610      2620      2630      2640      2650      2660      2670      2680      2690      2700
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00184CACGTGAATTCCGTGTTCTTTGGCTTGTTTGAGTATGCCAAGGAATGTTTCGTTTGTGAAAAGCCTGTTAATGATTACTTGGAATTTTGGAGACCTCAGA
00185.....
00187.....
00188.....
00189.....
00191.....
00193.....

```

```
00194 .....
00195 .....
00186 .....
00190 .....
00192 .....

      2710      2720      2730      2740      2750      2760      2770      2780      2790      2800
      |.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00184 CGTTCGACCATGCTCTTGAATCCGGTGATGATGTCATTTTTTCTTGTAACAGAGCCTTGAATTTCTCAACAGGAAGTACGGCTTTGACATCTGCTATAAG
00185 .....
00187 .....
00188 .....
00189 .....
00191 .....
00193 .....
00194 .....
00195 .....
00186 .....
00190 .....
00192 .....

      2810      2820      2830      2840      2850      2860      2870      2880      2890      2900
      |.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00184 ACCAGCAACACCCCGGTGATTCTTCCTAGAGCCATGCGAGGAACTAATGGGGGAAGACCCATGAATTCGTGAAGCTTGTTGACTACTTTATAAACATCC
00185 .....
00187 .....
00188 .....
00189 .....
00191 .....
00193 .....
00194 .....
00195 .....
00186 .....
00190 .....
00192 .....

      2910      2920      2930      2940      2950      2960      2970      2980      2990      3000
```

[illegible]

```

00193.....
00194.....
00195.....
00186.....
00190.....
00192.....

          3210      3220      3230      3240      3250      3260      3270      3280      3290      3300
          .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00184TTCAGAGTTTGGACGATGTGCTGCACCTTCTTTGTTGACAGCCAGTGATCTACGAACCTTTTGGAAATCGTGGAGATTAAAGAGTCTCAATGACATTTCTGA
00185.....
00187.....
00188.....
00189.....
00191.....
00193.....
00194.....
00195.....
00186.....
00190.....
00192.....

          3310      3320      3330      3340      3350      3360      3370      3380      3390      3400
          .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00184ACGCCTCTGAGTTCTTCAATTTATTGTAGTACATGGCTTTGATTTCTTGAAGTGGCAGGAGGTTCTTGATGTCTTTAACTAAGCCAGCCATACCCCCCGA
00185.....
00187.....
00188.....
00189.....
00191.....
00193.....
00194.....
00195.....
00186.....
00190.....
00192.....

```

```

          3410      3420      3430      3440      3450      3460      3470      3480      3490      3500
      .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00184GATGGCAATGCGTCTTGTTGGGTATCATGGCTGGTAGTCCAATCACCTCATTCACTTTGTTGATCAAATCAGTTACATCGAGCCCGAATCCTTTCAGGTAT
00185.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00187.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00188..A.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00189.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00191..A.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00193..A.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00194..A.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00195..A.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00186.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00190.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00192..A.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

          3510      3520      3530      3540      3550      3560      3570      3580      3590      3600
      .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00184TTGATGAATGCAACAACATCTGGGATGCTCTCAACTTGCTGAGTACCAAATTCCTTGAAGTCTTCGGTTTTTAATGTACTCTAATACTTTTTGCACCTCTTTGT
00185.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00187.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00188.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00189.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00191.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00193.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00194.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00195.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00186.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00190.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00192.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

          3610      3620      3630      3640      3650      3660      3670      3680      3690      3700
      .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00184CATTCCTGCAGGTACTCTTGAATGACAGGTATAATCTTCTCGAGGGGCAGAAGCGATAGAAAATCATCTACGTTCATCGGTAAGTGATCTACTCTCGTACAG
00185.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00187.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00188.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
00189.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

```

```

00191 .....
00193 .....
00194 .....
00195 .....
00186 .....
00190 .....
00192 .....

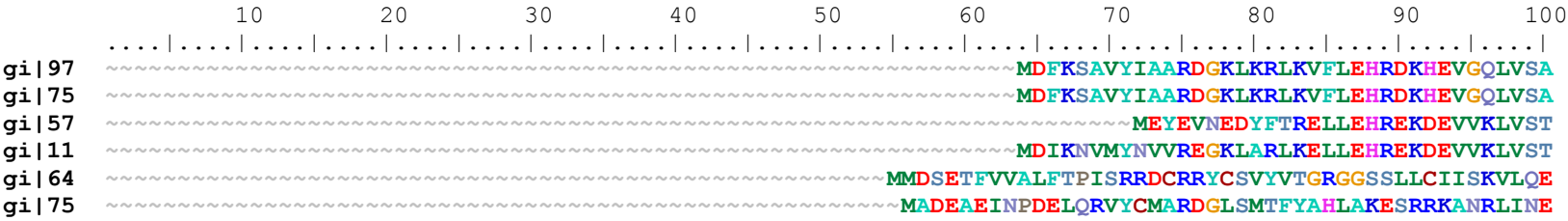
          3710      3720      3730      3740      3750      3760      3770      3780      3790
          .....|.....|.....|.....|.....|.....|.....|.....|.....|
00184AGGGATGCTGTTTATTGTGCTTGTCAAGCTGATCACCAGCAGGAAGGCCAGCGAGAACTTCATTGTGAAGTGGTTGAGCTGATGCTGATC
00185 .....
00187 .....
00188 .....
00189 .....
00191 .....
00193 .....
00194 .....
00195 .....
00186 .....
00190 .....
00192 .....

```

## ANEXO 7.

- Alineamiento general de secuencias de aminoácidos correspondientes a las variantes del gen fem. Se muestra en esta sección el alineamiento global entre las proteínas identificadas como variantes de *feminizer* en especies relacionadas y los isotigs pertenecientes a *D. longicaudata*

gi|970884209|ref|XP\_015126732.1| PREDICTED: protein fem-1 homolog CG6966 [Diachasma alloeum]  
gi|755935814|ref|XP\_011315450.1| PREDICTED: protein fem-1 homolog CG6966 [Fopius arisanus]  
gi|571534146|ref|XP\_006564184.1| PREDICTED: protein fem-1 homolog CG6966-like isoform X1 [Apis mellifera]  
gi|110757512|ref|XP\_392810.3| PREDICTED: protein fem-1 homolog CG6966-like isoform X2 [Apis mellifera]  
gi|645002558|ref|XP\_008207912.1| PREDICTED: feminization 1-like isoform X1 [Nasonia vitripennis]  
gi|755934447|ref|XP\_011314979.1| PREDICTED: protein fem-1 homolog B [Fopius arisanus]  
gi|970879625|ref|XP\_015127128.1| PREDICTED: protein fem-1 homolog B [Diachasma alloeum]  
Pfam12796 Ank\_2  
Pfam13637 Ank\_4  
Cdd:pfam13637 Ank\_4  
Cdd:cd00204  
mRNAIsotig06880 +3  
mRNAIsotig06880b +1  
Isotig06260 marco -2  
Isotig07202 marco-3





gi|97  
Pfam1  
Pfam1  
Cdd:p  
Cdd:c  
mRNAI  
mRNAI  
Isoti  
Isoti

MAGEAEIDQDLHRYVCMARDGLSMTLYAHLADKPREFVNRLINQ  
LSINKRRRSSSSDCNRVFKCRTKRIGVKCNKPWILVFSFSYSVDLWCMELLRLWTVQNRADKMDFKSAVYNAARDGKLKRLKVFLHRDKHEVGQLVSA  
MDKGALPSLQQLAMAMI

```

      210      220      230      240      250      260      270      280      290      300
...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|
gi|97  KFLVQH GADIEA ANRHGHTCLMIA CYRGH IRIAKY LLALKAD TNRSVK GNTALH DCAES GSLEIL KILLDH GARMEV DSYGMT PLLAAV AGRTHIVEH
gi|75  KFLVQH GADIEA ANRHGHTCLMIA CYRGH IRIAKY LLALKAD TNRSVK GNTALH DCAEAG SLEILK ILEHGA RMEVD SYGMT PLLAAV AGRTHIVEH
gi|57  RFLVTH GADIEM SNRHGHTSLMIA CYKGHI KIVKFL LALKAN VNRKSK GNTALH DCAES GSLEV VKVLLI EHGA RMGV DSYGMS PLLTAA VTGHKH IVEY
gi|11  RFLVTH GADIEM SNRHGHTSLMIA CYKGHI KIVKFL LALKAN VNRKSK GNTALH DCAES GSLEV VKVLLI EHGA RMGV DSYGMS PLLTAA VTGHKH IVEY
gi|64  KYLTDH SANIHI ANKHNTCLMIA AYKGHL DIVSFLE NGADP NEQAQY GETALH FFAEC GHCTIV KELLKH GSKMTK NAKNMS PLISAA ERARAD VVDC
gi|75  NYLVDH GGGDIH IANKYNTCLMIA AYKGHMD ILTFLE HGANP NEKALC GATALH FFAEC GHRHIV SKLLKY GTLVTK NVSGMT PLMAAA ERTRAN VVEC
gi|97  NYLVDH GADLHI ANKYNTCLMIA AYKGHL DIMTFLE HEANP NEKALC GATALH FFAEC GHCNIV SELLYK GATITK NVSGMT PLIAAA ERTRAN VVEC
Pfam1  ~~~~~
Pfam1  ~~~~~
Cdd:p  ~~~~~
Cdd:c  ~~~~~
mRNAI  ~~~~~
mRNAI  ~~~~~
Isoti  ~~~~~
Isoti  ~~~~~

      310      320      330      340      350      360      370      380      390      400
...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|
gi|97  LTGIPH LVNRKER IDALELL GATYVD KKRDMIL ~~ALEYWK RAMEERY RGDGPV IEKPPR PPPVAA YDYAQEV SEPEQL DELLADP DEMRMQ ALVIRERI
gi|75  LTGIPH LVNRKER IDALELL GATYVD KKRDMIL ~~ALEYWK RAMEERY RGDGPV IEKPPR PPPVAA YDYAQEV SEPERL DELLADP DEMRMQ ALVIRERI
gi|57  FISIPN LVSRKER IDALELL GATYVD KKRDMG ~~ALECWQ AMDERY RGDGPV IEKPPSP ~VVAAYD FAREIT DPDALN GLLNDP DEMRMQ ALVIRERI
gi|11  FISIPN LVSRKER IDALELL GATYVD KKRDMG ~~ALECWQ AMDERY RGDGPV IEKPPSP ~VVAAYD FAREIT DPDALN GLLNDP DEMRMQ ALVIRERI
gi|64  LIEHEDV ~SKEDMI EAYELL GASFAND KDHYCL QKAYTY LHKAMAL RYSDPNN IIKVPG EP~IAAYEN WREAE TVQRLE IIKNNH NAIHME SLTIRERI
gi|75  LISKNVA ~TKEEII DAYELL GASFAND KDNYCV ITAYTY LHKAMEM RYSDPSN IISKVVA EP~VPAYEN WRESE TLERLE SLMSNV NAIHME SLAIRERI
gi|97  LISENVA ~TKEEVI EAYELL GASFAND KDNYCL TTA YKYLHK AMKMRY SDPSNI NRKEVA EP~VHAYEN WRESE TLERLE SLKNDANAI HME SLAIRERI
Pfam1  ~~~~~
Pfam1  ~~~~~
Cdd:p  ~~~~~
Cdd:c  LLL~
mRNAI  ~~~~~
mRNAI  ~~~~~
Isoti  LISENVA ~TKEEVI EAYELL GASFAND KDNYCL STAYTY LHKAMEM RYSDSANI IR
Isoti  ~~~~~

```

gi|97 LGPAHPDTSYYIRYRGAIYADAGKFERCIELWNYALDMQOSMLEPLNPMTQ~SS~LFSFIELFSEFMVGEEGRQTGRGRRIPPVQREDLMRVFKKAVMEVK  
gi|75 LGPAHPDTSYYIRYRGAIYADAGKFERCIELWNYALDMQOSMLEPLNPMTQ~SS~LFSFIELFSEFMVGEEGRQTGRGRRIPPVQRADLMRVFTKAVMEV~  
gi|57 LGPAHPDTSYYIRYRGAVYADGGMFNRCIELWNYALDMQOSMLEPLDPMTQ~SS~LFSFTELFSEMI~~~GRQINTGRRVPPVQREELLRVFKKAVLEV~  
gi|11 LGPAHPDTSYYIRYRGAVYADGGMFNRCIELWNYALDMQOSMLEPLDPMTQ~SS~LFSFTELFSEMI~~~GRQINTGRRVPPVQREELLRVFKKAVLEV~  
gi|64 LGRQNPDVPQPVIYRGAVFADNARFDRCIDLWLHALNLRQ~~~~MNNMSI SKDLLRFQAQVSQMI~~~~~HVGVELKFLQ~~~VLNVLEASVVVELG  
gi|75 LGKHNPPEVPHPIVYRGAVLADNARFDRCIDLWLHALHLRQR~~~~NNITAV KDLLRFAQVFSQMI~~~~~HVGINQLQSQV~~~~MD~VLEASVVVELG  
gi|97 LGNHNPPEVPHPIVYRGAVLADNARFDRCIDLWLHALYLQR~~~~NSITAV KDLLRFAQVFSQMV~~~~~HVGIDILHFSQV~~~~MN~VLKASVIELG

Pfam1 ~~~~~~  
Pfam1 ~~~~~~  
Cdd:p ~~~~~~  
Cdd:c ~~~~~~  
mRNAI ~~~~~~  
mRNAI ~~~~~~ELFSFMVGEEGRQTVRGRRI PPVQREDLMRVFKKAVVEI~  
Isoti ~~~~~~  
Isoti ~~~~~~

gi|97 ~LGKQML~DKVPNCDRDLTNLNRI~LIITLHLACLTTREMP~~~~EVGSEE~YV~NLHRAIH ELVRINAKDKGERDVLQMTHNVDASIAGS~~~Y  
gi|75 RLKQKML~DKVPKCDRLTNLNRI~LIITLHLACLTTREMP~~~~QVGSEE~YV~KLHRAIH ELVRINAKDKGERDVLQMTHNADASITGS~~~Y  
gi|57 KLKQKMM~DKGPTRGRDIVYLDKV~LLTTLHLIASLLTHEMP~~~~EKDTAE~YT~ALHQALYELVRINAKDRNGGNVLHLVFRERHI VLGAGPKS  
gi|11 KLKQKMM~DKGPTRGRDIVYLDKV~LLTTLHLIASLLTHEMP~~~~EKDTAE~YT~ALHQALYELVRINAKDRNGGNVLHLVFRERHI VLGAGPKS  
gi|64 RNKKKLI~EPEPKEDSQSISEEIESNI TITLYILTILTKMLTN~EKG YDETDIK~RAHHLVHKLCALQVTLKDSQTLLHLAVNA DTPVDDF~~~H  
gi|75 RNKTRMLASEAKEDDEQYLEETESN~VM TTYIILTILVQLMP~LKG KSYDEP~SISKAREFLVHKLCALKMTLRDQS QTLLHLAVNAKTPVDDF~~~H  
gi|97 RNKTRMLASDAKEDKEQSILEETESN~VM TTYIILTILVKLMA~INEKACDED~TVEKTTFIVRKLCALKMTLRDQS QTLLHLAVNAKTPVDDF~~~H

Pfam1 ~~~~~~  
Pfam1 ~~~~~~  
Cdd:p ~~~~~~  
Cdd:c ~~~~~~  
mRNAI ~~~~~~  
mRNAI ~~~~~~KLKQKML~DKIPNCDRDLTNLNRI~LIITLHLACLTTREMP~~~~EVGSEE~YV~NLHRAIYELVRINAKDKGERDVLQMIHNVEASIAGS~~~Y  
Isoti ~~~~~~  
Isoti ~~~~~~

	610	620	630	640	650	660	670	680	690	700	
	..... ..... ..... ..... ..... ..... ..... ..... ..... ..... .....										
gi 97	P~~ACKLTSPQLTTA~~	LLRAGSDPTATDDEGNTALHLTAMSHPWRTDLATL	~~~~~	LLKAGAHIDAVNNEGQTYQSLLDDKAKYDSLNP	~~~~~	LLKAGAHIDAVNNEGQTYQSLLDDKAKYDSLNP	~~~~~	LLKAGAHIDAVNNEGQTYQSLLDDKAKYDSLNP	~~~~~	LLKAGAHIDAVNNEGQTYQSLLDDKAKYDSLNP	
gi 75	P~~GCKLTSTQLTTA~~	LLRAGSDPTATDDEGNTALHLTAMSHPWRTDLATV	~~~~~	LLKAGAHIDAVNNEGQTYQTLDDKERYDSLNP	~~~~~	LLKAGAHIDAVNNEGQTYQTLDDKERYDSLNP	~~~~~	LLKAGAHIDAVNNEGQTYQTLDDKERYDSLNP	~~~~~	LLKAGAHIDAVNNEGQTYQTLDDKERYDSLNP	
gi 57	P~~TYRFPSPNLIKA~~	LIRVGADVTATDMTDNTVLHLAAY~HYPSLDLFTIL	~~~~~	LDAGAHIDAVNKSQDTFEKLTWRKRPYDAVYLVKYTTLACL	~~~~~	LDAGAHIDAVNKSQDTFEKLTWRKRPYDAVYLVKYTTLACL	~~~~~	LDAGAHIDAVNKSQDTFEKLTWRKRPYDAVYLVKYTTLACL	~~~~~	LDAGAHIDAVNKSQDTFEKLTWRKRPYDAVYLVKYTTLACL	
gi 11	P~~TYRFPSPNLIKA~~	LIRVGADVTATDMTDNTVLHLAAY~HYPSLDLFTIL	~~~~~	LDAGAHIDAVNKSQDTFEKLTWRKRPYDAVYLVKYTTLACL	~~~~~	LDAGAHIDAVNKSQDTFEKLTWRKRPYDAVYLVKYTTLACL	~~~~~	LDAGAHIDAVNKSQDTFEKLTWRKRPYDAVYLVKYTTLACL	~~~~~	LDAGAHIDAVNKSQDTFEKLTWRKRPYDAVYLVKYTTLACL	
gi 64	TNDVCKFPKCAATAKL~~	LIQCGADVAMDNERNTPHLVIVQYKPISEDFMTLH	~~~~~	SIILDLVEAGAHMDTVNSQKTPFDAATTGVAEII	~~~~~	SIILDLVEAGAHMDTVNSQKTPFDAATTGVAEII	~~~~~	SIILDLVEAGAHMDTVNSQKTPFDAATTGVAEII	~~~~~	SIILDLVEAGAHMDTVNSQKTPFDAATTGVAEII	
gi 75	T~~NDVCKFPKCAATARLLIECGADVAMDNERNTPHLVIVTYTKPISEDFMTLH	~~~~~	GIILELIEAGAHMDIVNSQGETPYDAATTGVAEII	~~~~~	GIILELIEAGAHMDIVNSQGETPYDAATTGVAEII	~~~~~	GIILELIEAGAHMDIVNSQGETPYDAATTGVAEII	~~~~~	GIILELIEAGAHMDIVNSQGETPYDAATTGVAEII	~~~~~	GIILELIEAGAHMDIVNSQGETPYDAATTGVAEII
gi 97	T~~NNVCKFPCLATAQLLIRGGADVAMDNERNTPHLVIATYQKPISEDFMTLH	~~~~~	SIILELIDAGAHMDIVNSQGETPYDAATTGVAEII	~~~~~	SIILELIDAGAHMDIVNSQGETPYDAATTGVAEII	~~~~~	SIILELIDAGAHMDIVNSQGETPYDAATTGVAEII	~~~~~	SIILELIDAGAHMDIVNSQGETPYDAATTGVAEII	~~~~~	SIILELIDAGAHMDIVNSQGETPYDAATTGVAEII
Pfam1	~~~~~	EIVKL~~	LENGADVNAKDKGNTALHLAARN~~	GNLEIVKL	~~~~~	LEHGA~~	D~INLKD	~~~~~			
Pfam1											
Cdd:p											
Cdd:c											
mRNAI											
mRNAI	P~~VCKLTSPQLTTA~~	LLRAGADPTATDDEGNTALHLTAMSHPWRTDLATL	~~~~~	LLRAGAHIDAVNNEGQTYQRLNDKAKYDSLNP	~~~~~	LLRAGAHIDAVNNEGQTYQRLNDKAKYDSLNP	~~~~~	LLRAGAHIDAVNNEGQTYQRLNDKAKYDSLNP	~~~~~	LLRAGAHIDAVNNEGQTYQRLNDKAKYDSLNP	
Isoti											
Isoti	~~~~~									QTKLSLQCM	
	710	720	730	740	750	760	770	780	790	800	
	..... ..... ..... ..... ..... ..... ..... ..... ..... ..... .....										
gi 97	AARVIRR~TAIH~INRVPWHLREFVTQH										
gi 75	AARVIRR~TANILINRVPWHLHQFVIQH										
gi 57	AARVVRK~TYDISFVPKNLQDFVLMH										
gi 11	AARVVRK~TYDISFVPKNLQDFVLMH										
gi 64	AAKAVKAYNLLYCGNVPRSLESFIELHGPGLNQG										
gi 75	AAKAVKTYNLSYSGNVPRSLESFIELHGSGLHQG										
gi 97	AAKAVKTYNLSYSGNVPRSLESFIELHGSGLHQG										
Pfam1											
Pfam1											
Cdd:p											
Cdd:c											
mRNAI											
mRNAI	AARVIMR~T										
Isoti											
Isoti	AAKAVKTYNLSYSGNVPRSLESFIELHGPGLHQG	PLPRRFVD	DEDDIFFDILFAKT	VAVWYVEFSMQC	ILSSNFIFIFICCLHYRL	FIVQTFV	KFIRYDFW				

```

      810      820
      ....|....|....|....|.
gi|97
gi|75
gi|57
gi|11
gi|64
gi|75
gi|97
Pfam1
Pfam1
Cdd:p
Cdd:c
mRNAI
mRNAI
Isoti
Isoti  VPLDIVPGSAYALVPFISLNL

```

## ANEXO 8.

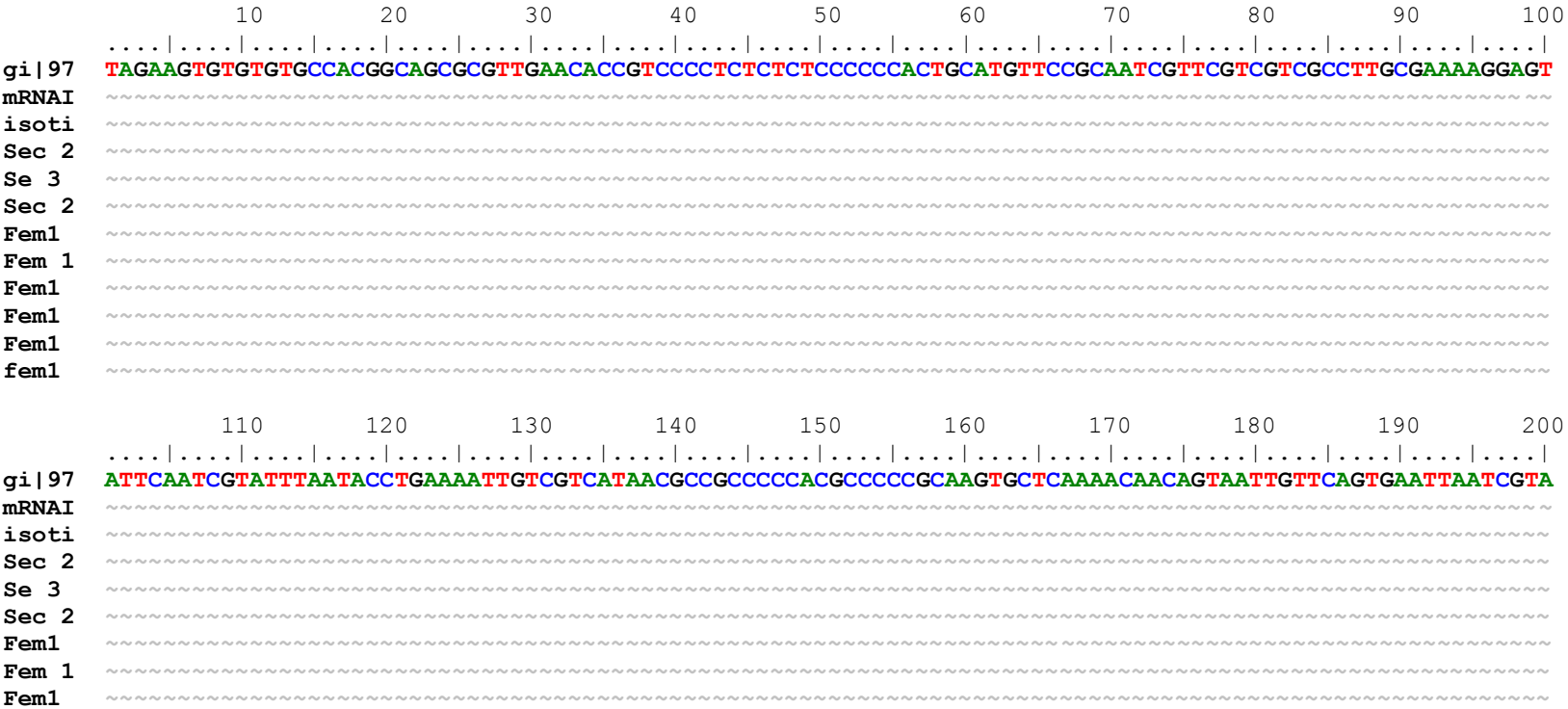
- Alineamiento de Isotigs 06880/06880b con la secuencia de *D. alloeum*, con la que mayor homología de secuencia presentaron en BLASTn. Se muestran las variantes obtenidas por medio de PCR de punto final para completar el transcripto.

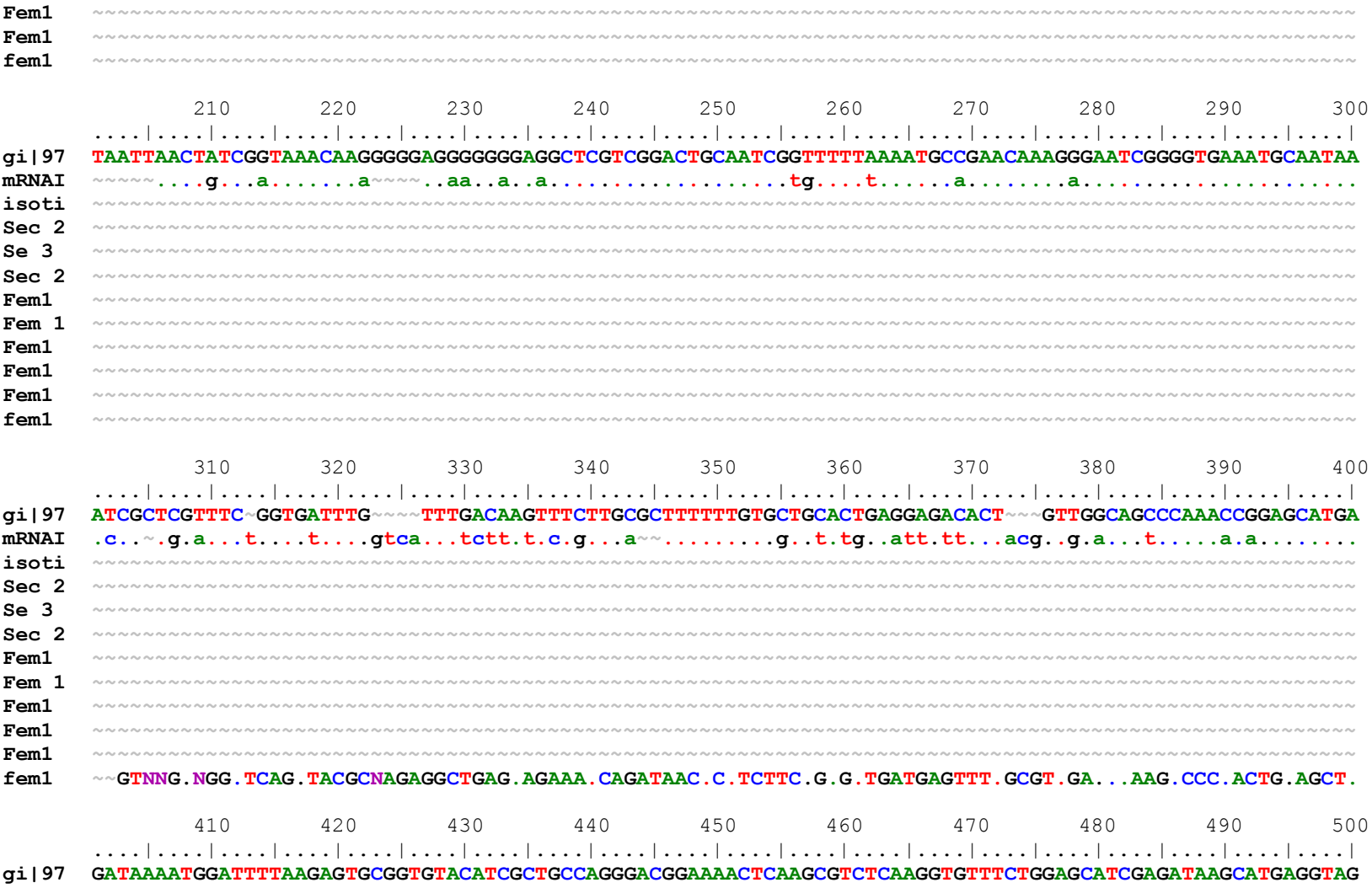
```

gi|970884208|ref|XM_015271246.1| PREDICTED: Diachasma alloeum protein fem-1 homolog CG6966 (LOC107048182), mRNA
mRNAIsotig06880
isotig06880bis mRNA
Sec 2 femCG
Se 3 femCG
Sec 23

```

Fem1 Dlongi F3-R1 Hem670 fw  
Fem 1 Dlongi F3-R1 Hem670 rev INV  
Fem1 Dlongi F3-R1 Hem700 fw  
Fem1 Dlongi F3-R1 Mac1000 rv  
Fem1 Dlongi F3-R1 Hem700 rev  
fem1 Dlongi F3-R1 Mac1000 fw





mRNAI .....a.....  
isoti  
Sec 2  
Se 3  
Sec 2  
Fem1  
Fem 1  
Fem1  
Fem1  
Fem1  
fem1 TCC.GCCG.C..AA.CCTCCAT.CAGTA..GGT.GCA.G.TTTCTTCCCCTTCAG.T.AG.GAG.CTAT.GAT...CT..TCTTC.TGA..ACC.CA.C

510 520 530 540 550 560 570 580 590 600  
gi|97  
mRNAI .....t.....a.....t..  
isoti  
Sec 2  
Se 3  
Sec 2  
Fem1  
Fem 1  
Fem1  
Fem1  
Fem1  
fem1 CG.TGG.TT.TC.TTGT.TTTAGTC..T.C.GTTC.G.AGCTC.CG.AAACAAT..GT.TA...C.GAA..CATCAACT.CC.G.TTGAC.CTC.AT.CC

610 620 630 640 650 660 670 680 690 700  
gi|97  
mRNAI GGCTGATATTGAACAGCCTGGATCAGTTGTCTTTGACGGTGAGACGATAGAGGGTGACCTCCACTCTGGTGTGCAGCGGCTGCTGGTTCATCTAGCTCTT  
isoti .....g...g...a.....t.....a.c.....t.....t.....c.g..t.t.....  
Sec 2  
Se 3 .....G..A.G...A...N.....T.....C.....T.....T.....C..G..T.T.....  
Sec 2  
Fem1  
Fem 1



```
Fem1
Fem1
Fem1
fem1 TCAATGAGGGTG.A.T.AGCTGCATCA.CC..CATTGCT.CCA.AC.CTC.ACTCTACTA.TTCTCT..ACACA.TCGAT...GG.TCTCG.ATCTCT.G

      710      720      730      740      750      760      770      780      790      800
gi|97  ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
mRNAI GTCAAATTGCTCGTCAAACGGGGAGCGAGATTGAATTCAACGACAACGACAAATTCAACGCCTCTGAGGGCGGCTTGCTT~CGATGGACACTTCGAAATT
isoti  .....a..t....g..c.....a.....c..t..g.....c.....t..t.....
Sec 2  ....
Se 3   .....A..T....G..C....A.....C..T..G.....T..T.....
Sec 2  .....NNNNC..T..G.....T..T.....
Fem1   ....
Fem 1  ....
Fem1   ....
Fem1   ....NN.....T..T.....
Fem1   ....
fem1   AAT..GCA...TTCGCC.GC.CCTCAC.CT.CTT.GGACT.TTGCTCTCC.C..ACGC..CAACGACCC.TTCTT.CCACAG.TCCA..GTTCTCTCGA.

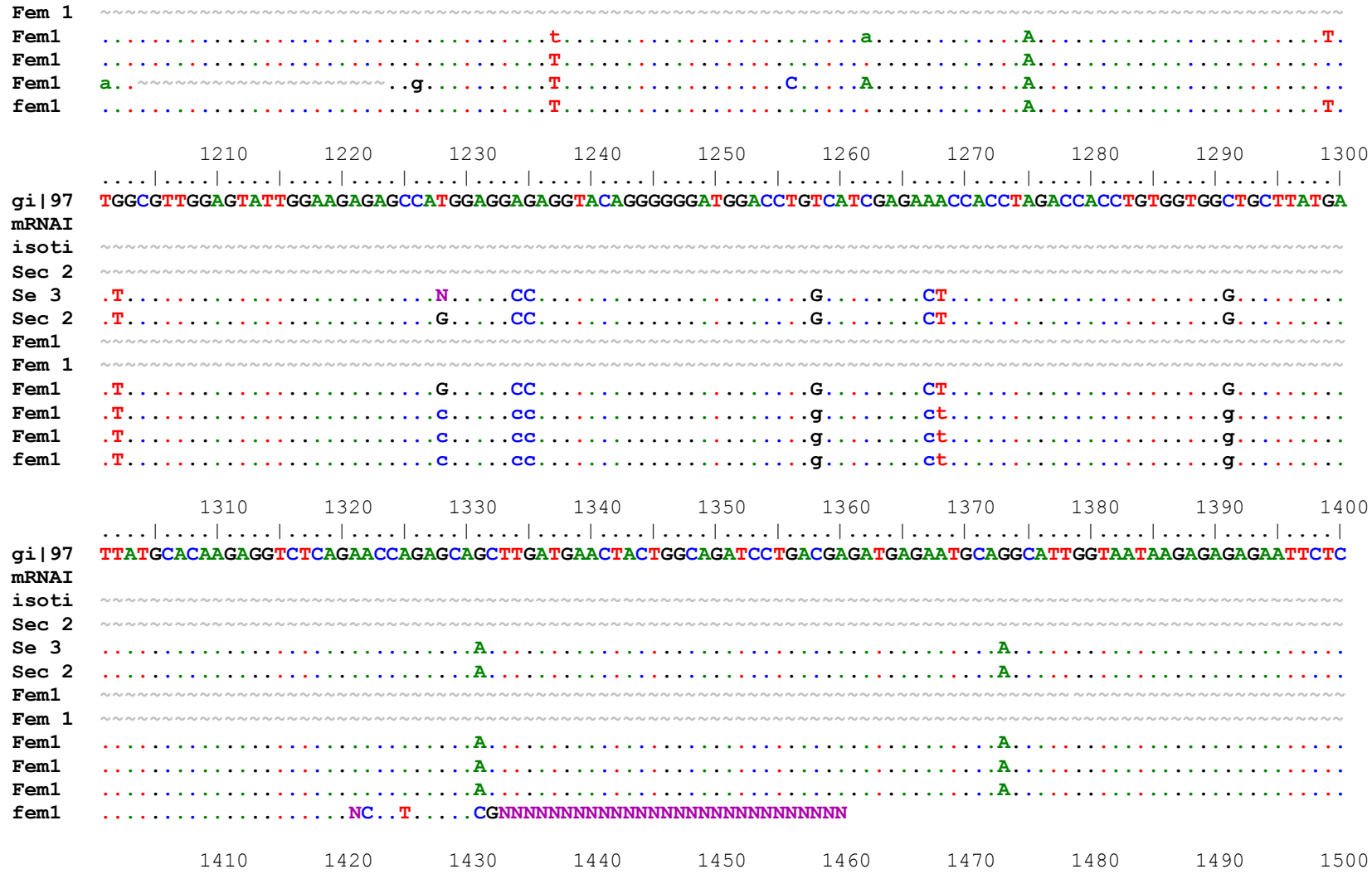
      810      820      830      840      850      860      870      880      890      900
gi|97  ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
mRNAI GTTAAATTTCTGGTTCAGCATGGGGCTGACATCGAAGCAGCGAATCGTCACGGTCACACCTGCCTGATGATCGCCTGCTACCGTGGCCACATCAGAATAG
isoti  .....a.....a.....c....a....c.....t.....a..t....t....t.....a.
Sec 2  ....
Se 3   .....A.....A.....C....A....C.....T.....A..T...T.C...T.....A....
Sec 2  .....A.....A.....C....A....C.....T.....A..T..A...C...T.....A....
Fem1   ....
Fem 1  ....
Fem1   ....
Fem1   .....A.....A.....C....A....C.....T.....A..T..A...C...T.....A....
Fem1   ....
fem1   TACGTCCAAG.AC..AGCAT.C.TCAG.TGT..A.TTGGTGAT.CT.GGG.T...TTCA.G...A.GGTC.TAT.GT.C.AGAG.AGA.TG.TCTTCC.C

      910      920      930      940      950      960      970      980      990     1000
.....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
```

```
gi|97 CCAAGTACCTCCTAGCCCTGAAAGCCGACACCAACCGTAAATCCGTCAAAGGTAATACAGCTCTCCACGACTGCGCAGAGAGTGGCTCCCTGGAGATATT
mRNAI
isoti
Sec 2
Se 3 .....A.....C.....G.....G.....A.....
Sec 2 .....A.....A.....C.....G.....G.....A.....
Fem1
Fem 1
Fem1 .....A.....A.....C.....G.....G.....A.....
Fem1 .....C.....G.....G.....A.....
fem1 .A.TT.G.A.GT.CCTTTT.C.CC.A.TT...G.GTTCTGTCAAT..CTCCTGG..GGCAT.GA.AT.CTTAA.CTTTCCGT.A.TCATA.ATCCC.GCC

      1010      1020      1030      1040      1050      1060      1070      1080      1090      1100
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
gi|97 GAAGATACTCCTGGACCATGGGGCCCGGATGGAGGTGGACTCCTACGGAATGACTCCTCTCCTCGCAGCGGCTGTGGCTGGTCCGACCCACATCGTCGAG
mRNAI
isoti
Sec 2
Se 3 .....C.....G.....A.....CCG.....t.....
Sec 2 .....C.....G.....A.....CCG.....T.....
Fem1
Fem 1 .....C.....G.....A.....CCG.....T.....
Fem1 .....C.....G.....A.....CCG.....T.....
Fem1 .....C.....G.....cgg.....t.....a
fem1 AGG.TCTGG.AGCCCA.G..CAAGGTTTGGT.GTTGCAT.GATGGGAAC.TGG.G.CA..AGGA..AAATACCT.TAACCAC TTGGGT.G.....

      1110      1120      1130      1140      1150      1160      1170      1180      1190      1200
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
gi|97 CATCTCATCGGGATTCTCATCTGGTGAATCGTAAGGAGAGGATCGATGCTCTTGAAC TCCTGGGGCGACGTATGTCGACAAGAAACGAGATATGATCC
mRNAI
isoti
Sec 2
Se 3 ...T.....N.....A.....A.....
Sec 2 .....T.....A.....
Fem1
```



gi|97  
mRNAI  
isoti  
Sec 2  
Se 3  
Sec 2  
Fem1  
Fem 1  
Fem1  
Fem1  
Fem1  
fem1

GGTCCGTGCACATCCAGACACTAGTTACTACATAAGGTATCGTGGGGCTATCTATGCTGATGCTGGCAAGTTTGAGAGGTGCATTGAACCTGTGGAATTACG

1510 1520 1530 1540 1550 1560 1570 1580 1590 1600

gi|97  
mRNAI  
isoti  
Sec 2  
Se 3  
Sec 2  
Fem1  
Fem 1  
Fem1  
Fem1  
Fem1  
fem1

CACTGGATATGCAGCAGAGCATGCTGGAGCCGTTGAATCCCATGACCCAGAGCTCGTTGTTTCAGTTTTATTGAGTTATTTAGTTTTATGGTTGGTGAGGA

1610 1620 1630 1640 1650 1660 1670 1680 1690 1700

gi|97  
mRNAI  
isoti  
Sec 2  
Se 3  
Sec 2

GGGGAGACAGACGGGGAGGGGGAGGAGAAATACCACCGTGCAGAGGGAGGATCTTATGAGGGTGTTTAAAAAGCTGTTATGGAAGTTAAACTGGGAAAG

a..a.....t..a.....t..a....a....a.....t..c.....g..g....a.....a..c..

A.....T..ANNNNNNNN.....NNNN

Fem1  
Fem 1  
Fem1 A . . . . . T . A . . . . . T . A . . . A . A . . . . . T . C . . . . . G . . G . . . A . . . . A . C . .  
Fem1 A . . . . . T . A . . . . . T . A . . . A . A . . . . . T . C . . . . . G . . G . . . A . . . . A . C . .  
Fem1  
fem1

gi | 97 CAGATGTTGGACAAAGGTACCCAACTGTGACCGTGACCTCACGAAATTTAAACAGGATACCTTATTAATCACTCTTCATCTCGCCTGTCTGTGGACGAGAGAA

isoti . . . . . a . . . . . c . . . . . t . . . . . g . . . . . c . . . . . t . . . . . t . a . a . . . . .

Se 3

**Fem1** ... T . A . A . . . .

Fem1 . . . . . A . . . . . C . . . . . T . . . . . G . . . . . C . . . . . T . . . . . T . a . A . . . . .

	Female	Male
Age	10.0	10.0
Sex	Female	Male
Height	1.60	1.70
Weight	55.0	70.0
Body Mass Index	21.5	24.0
Heart Rate	70	75
Blood Pressure	120/80	130/90
Cholesterol	180	200
Blood Sugar	90	95
Glucose Tolerance	100	110
Insulin Sensitivity	1.0	1.0
Metabolic Rate	1500	1800
Energy Expenditure	2000	2500
Physical Activity	30	40
Stress Level	50	60
Sleep Quality	70	80
Mood Stability	80	90
Emotional Resilience	90	100
Overall Health	100	100

1910 1920 1930 1940 1950 1960 1970 1980 1990 2000

si | 97 TGGCGGGAAGTTGGGTCTGAGGAGTATGTCGAATCTTCATCGGGCTATCCAGGAGTTGGTTCTGAATTAAATGCCAAAGATAAAGGAGGACGAGATGTTCTTCCA

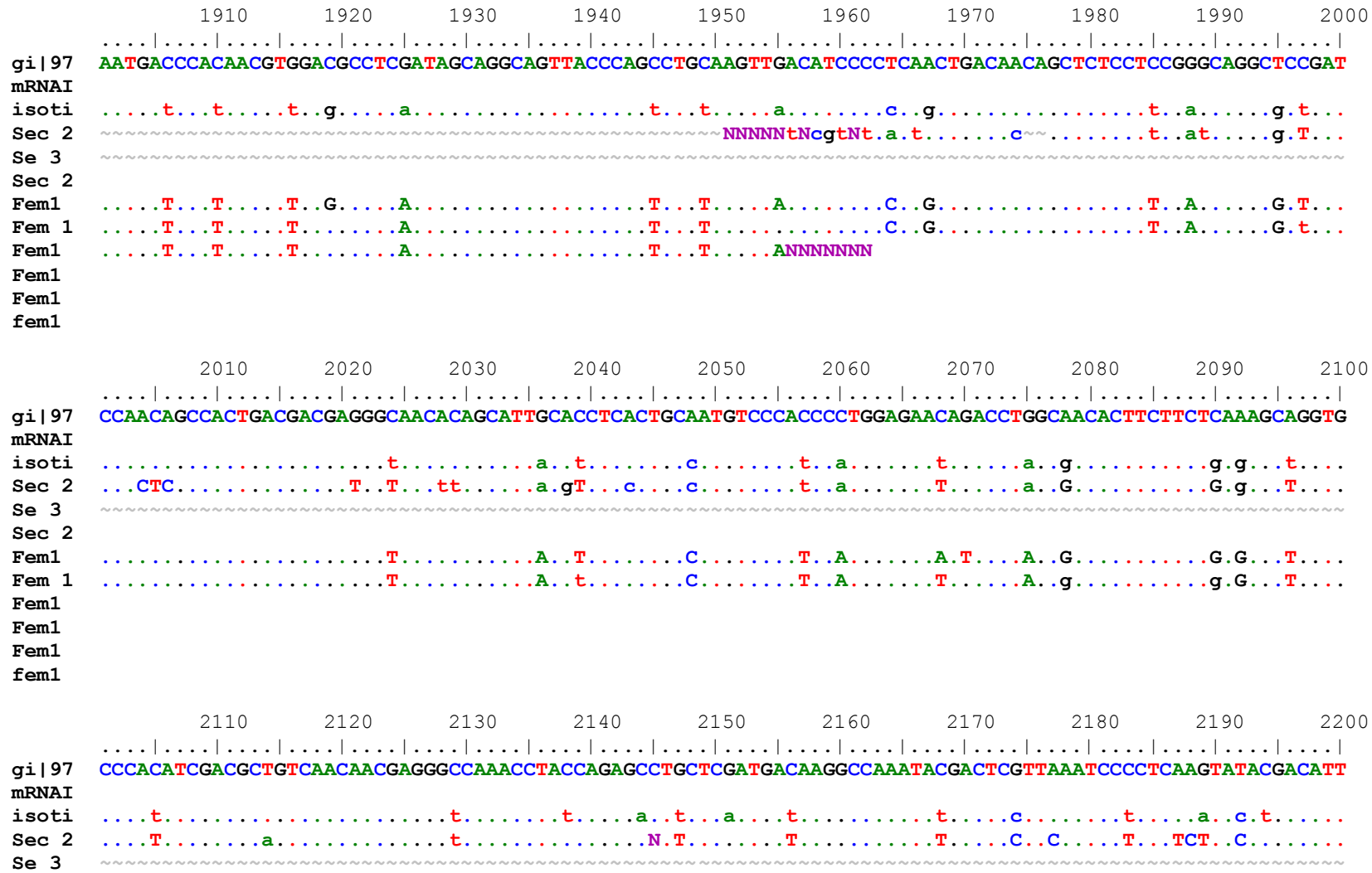
isoti . . . a . . . a . . . a . . . t . . . a . . . ca . . . t . . .

Se 3

Fem1 . . . A . . . . . A . . . . . A . . . . . T . . . . . A . . . . . GC . . . . . CG . . . . . T . . . . .

Fem1 . . . A . . . . . A . . . . . A . . . . . T . . . . . A . . . . . GC . . . . . CG . . . . . T . . . . .

Fem1



Sec 2  
Fem1 .....T.....A.....T.....T.....A..T..A.....T.....T.....C.....T.....A..C.T.....  
Fem 1 .....T.....A.....T.....T.....A..T..A.....T.....T.....C.....T.....A..C.T.....  
Fem1  
Fem1  
Fem1  
fem1

2210 2220 2230 2240 2250 2260 2270 2280 2290 2300  
gi|97 .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|  
mRNAI GAAGTGCTTAGCTGCAAGAGTCATCAGGAGAACAGCCATTACATCAACAGGGTACCCCTGGCATCTTCGTGAATTTGTCAACCAACACTAGTTCTCGAAT  
isoti .c.....a..cc...t...t.....  
Sec 2 .C.....A..CC...T...T.....t..CTTG...a.ac.....C.....C.....CGTCT.  
Se 3 ~~~~~  
Sec 2  
Fem1 .C.....A..CC..N.T...T.....T..CTTG...A.AC.....C.....C.....CGTCT.  
Fem 1 .C.....A..CC...T...T.....t..CTTg...A.a.....C.....C.....CGTCT.  
Fem1  
Fem1  
Fem1  
fem1

2310 2320 2330 2340 2350 2360 2370 2380 2390 2400  
gi|97 .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|  
mRNAI GCTCGTGCGGATTCACTCTAGTCCGACTGCTCTAAATAAATCGTCGCCTCTCTTTTTTTTATACCTTCTCTGTTTATTTTTTGGCAGCCATACTTATCTG  
isoti  
Sec 2 .....CTGGTG.....G.....GGCC.GAGAGCTGGTGCCcATA.CCAANANGAANNATANCCGTGT  
Se 3 ~~~~~  
Sec 2  
Fem1 .....CTGGTG.....G.....GGCC.GAGAGCTGGTG.CCATA.CCAA.CAT.T...GG.A..A..  
Fem 1 ...NNNtNN  
Fem1  
Fem1  
Fem1  
fem1

```

                2410      2420      2430      2440      2450      2460      2470      2480      2490      2500
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
gi|97      TGTAAATAAGCTATGATGATGATGATGA
mRNAI
isoti
Sec 2      AA.TCC...NN.Ngtg..gNt.atNaNaNaactggNcNaNaNaNNNcNTTTTTTtcTcatttgtttTaNcattNtTataTcgTactgcacaaatTTaTct
Se 3      ~~~~~~TACGTTTTTAATTTGATTTATTTTGACTTTTTTTTATTTTTTTTTTTACGGGGTTTTTGGATTTTatgTTTTg
Sec 2
Fem1      AT.C.GT.C.ACT.A..CCAA.ACTA.TGTGATGGAAATCCCAACGAGAAGGCACCTCTGTANTTTTTTTTTTATANCCCNNGCNNNGTTTTGNCNTTTTTT
Fem 1
Fem1
Fem1
Fem1
fem1

                2510      2520
      ....|....|....|....|...
gi|97
mRNAI
isoti
Sec 2      gTTattcatTTtctGTTTcacca
Se 3      gggNTTaNNTTATNTTTTTTN
Sec 2
Fem1      TNNNNNG
Fem 1
Fem1
Fem1
Fem1
fem1
```

ANEXO 9.



- Alineamiento de Isotigs 07202/06260 con la secuencia de *F. arisanus* con la que mejor Hit de BLAST tuvieron. Se muestran las variantes obtenidas por medio de PCR de punto final para completar el transcripto.

gi|755934446|ref|XM\_011316677.1| PREDICTED: Fopius arisanus protein fem-1 homolog B (LOC105273937), mRNA

Isotigs07202/06260 fem1

Fem1 Dlongi F10-R13 H

FEm1 Dlongi H sec 20

Fem1 Dlongi sec8 M

FEm1 Dlongi M sec 21

Fem1 Dlonf F9R8 f H

Fem1 Dlongi F9-R8 rev INV H

Fem1 Dlongi sec13 M

Fem1 Dlongi sec14 M

Fem1 Dlongi sec15

Fem1 Dlongi sec 3

Fem1 Dlongi F1R1

Fem1 Dlongi F1R1rev invertida

Fem1 Dlongi F3-R3 for

Fem1 Dlongi F3-R3 rev invertida

	10	20	30	40	50	60	70	80	90	100
gi 75	....	....	....	....	....	....	....	....	....	....
Isoti	GGCGAT	CCCC	TAAAAA	GAGT	GATGG	CAGAT	GAGGCT	GAAAT	CAATCC	AGATGA
Fem1	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
FEm1	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
FEm1	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
FEm1	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Fem1	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Fem1	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Fem1	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
Fem1	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....

[illegible]

Sequence logo for the 75th position of the Isotripartite protein. The logo shows the conservation of amino acids at this position across 1000 randomizations. The x-axis represents positions 310 to 400, and the y-axis represents the probability of each amino acid. The sequence TATTGAAGGAGCTAGTGCCTTGTGGTGTGCTGCAGGCGCTGGACATCTTAATGTAGTTAAACTCTGGTGAAGCCGGAGCGGACGTGAATCAGCTACA is highlighted in green. The sequence C..c.....t.....a....g...t.g..c..ta.....a..c..c..t..t..t..c..t..C..C....A..G is highlighted in red. The sequence CATACAAATTCAACTCCACTCAGGGCGGCGTGTTCGATGGAAGGTTGGATATAGTCAACTACCTTGTAGACCACGGAGGCGATATACACATAGCTAATA is highlighted in blue. The sequence .....c.....t.....a..a.....a....c.....c..a..t..t....ca.t....g.ct...t.....c.... is highlighted in yellow.

[illegible]

[illegible][illegible]

Sequence logo for the 5' region of the 18S rRNA gene. The y-axis lists 15 sequences, with the first 10 labeled 'Fem1' and the last 5 labeled 'Isoti'. The x-axis shows nucleotide positions from 910 to 1100. The logo displays the conservation of nucleotides across these sequences, with color-coded bases (A, C, G, T) and a scale from 0 to 1.0. A prominent 'C' is seen at position 910, and a 'G' at 911. The sequence 'CGATCCGTCGAATATAATTAGCAAAGTAGTGGCAGAACCCAGTGCCCTGCATACGAGAATTGGAGAGAGAGTGAGACTTTGAAAGACTTGAGAGCTTGATG' is shown at the bottom, corresponding to the 'Isoti' sequence.

[illegible]

[illegible]



[illegible][illegible]

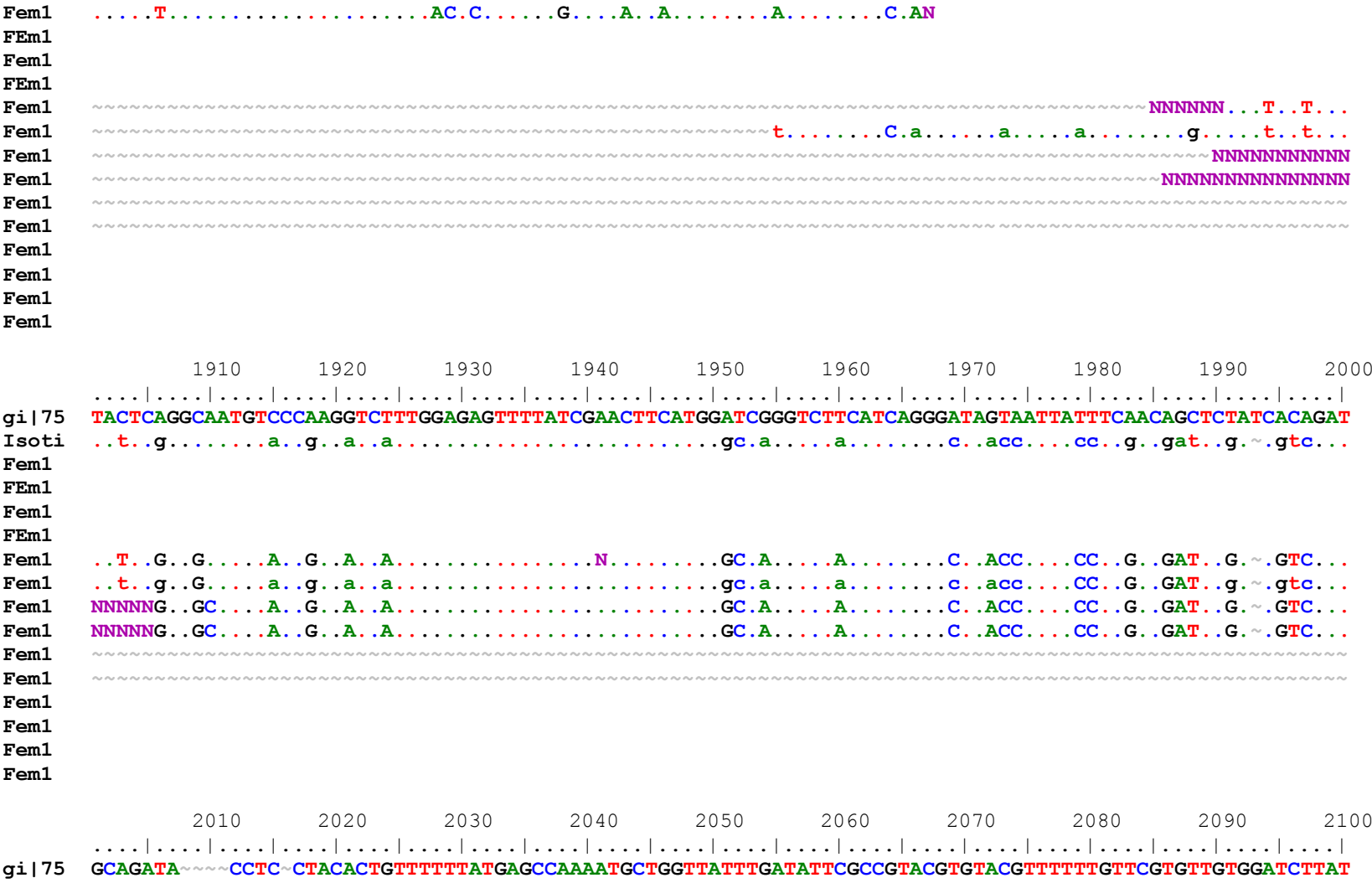
Fem1 T...C.N  
Fem1 C.NANNGANANNNNNNNCA.TT..TTT.T.TTTTT.TN  
FEm1  
Fem1  
Fem1  
Fem1  
Fem1  
Fem1  
Fem1  
Fem1  
Fem1  
Fem1  
Fem1  
Fem1

1710 1720 1730 1740 1750 1760 1770 1780 1790 1800

gi|75 CTATCAGCGATTATGACGCTTCATGGGATTATATTGGAGTTGATCGAGGCCTGGGGCCACACATGGACATTGTCAATAGCCAGGGAGAGACGCCGTATGA  
Isoti  
Fem1 AAT.AGTGAT...ATGAC.CTGCAACA.CATC.TAC...AGC....GAT.CTG....CATATG.A..TG.CAT...CAGA....C..C.TA.GA.G.T.  
Fem1  
Fem1  
FEm1  
Fem1  
Fem1  
Fem1  
Fem1  
Fem1  
Fem1  
Fem1  
Fem1  
Fem1  
Fem1  
Fem1

1810 1820 1830 1840 1850 1860 1870 1880 1890 1900

gi|75 TGCTGCAACAACGGGAGTTACTGAGATTATACTTAGAACACAGACTAAATTGTCCCTTCAGTGATGGCTGCCAAGCTGTCAAGACATACACCTCTCT  
Isoti  
.a.a.....a.....c.....a.....a.....g.....t.t.



```
Isoti .a.....tatt.t..ga..t...~.....a.....ccg...c.g....g...gt.ga..t.agtatgcaa.g.ata.t..aa.aatct...a..
Fem1
Fem1
Fem1
Fem1
Fem1 .A.....TATT.T..GA..T...~.....A.....CCG...C.G....G...GT.GA..T.AGTATGCAA.G.ATA.T..AA.AATCT...A..
Fem1 .a.....tatt.t..ga..t...~.....a.....ccg...c.g....g...gt.ga..t.agtatgcaa.g.ata.t..aa.aatct...a..
Fem1 .A.....TATT.T..GA..T...~.....A.....CCG...C.G....G...GT.GATGc.Ng...aaacG.GG.acGa..C..Ccacaac.G.
Fem1 .A.....TATT.T..GA..T...~.....A.....CCG...C.G....G...GT.GATGc.Ng...AAaCg.gg.ACgA..C..CCACAAC.g.
Fem1 ~~~~~
Fem1 ~~~~~
Fem1
Fem1
Fem1
Fem1

          2110      2120      2130      2140      2150      2160      2170      2180      2190      2200
gi|75      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Isoti      CAGCTTCTCGCAATGACAACCGTACATTACGCGGCCTATAACGCTCACCTATGACAAAAGGGCTTTCGTAGCTATCGAAGTAGCTACTTGCCCATGTGTT
Fem1      t.at..a.tttc..atth.tttgtt.a.g.tt.ca....cg.tga.tgt.cat.gtgc..act...gta.aa.t.at.c..ta.g....ctggg.accac
Fem1
Fem1
Fem1
Fem1      T.AT..A.TTTC..ATTT.TTTGTT.A.G.TT.CA....CG.TGA.TGT.CAT.GTGC..ACT...GTA.AA.T.AT.C..TA.G....CTGGG.ACCAC
Fem1      t.at..a.tttc..atth.tttgtt.a.g.tt.ca....cg.tga.tgt.cat.gtgc..act...gta.aa.t.at.c..ta.g....ctggg.accac
Fem1      TN..gcA.Aa.CC.aTtCgg.TGGTTGG.TTG..T..NNTNN.NGANG
Fem1      tN..Gca.aA.cC.atTCgG.tGgttgg.ttg..t..Nacta..catatcccc.ttttg.tcaa..gN.tNNg.g.TTC..G..GGT...TTAGGAgc..
Fem1 ~~~~~
Fem1 ~~~~~
Fem1
Fem1
Fem1
Fem1

          2210      2220      2230      2240      2250      2260      2270      2280      2290      2300
          ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
```

```

gi|75 GTAAATTGGTATCTGCAGGATTGGCTACGAGAAGAGTTTGTTTTACAGTGAGAGTAAACTGGTCAAATGATGAAGAAAACAGAAATTGTGGATTTGCCC
Isoti tgg.ca.t..tc.a.gca.c.aa..atatgccttg..acca.....a.t.tcattha..tct.tta.cttcagg...atgcttttctact.ttg.aattga
Fem1
FEm1
Fem1
FEm1
Fem1 TGG.CA.T..TC.A.GCA.C.AA..ATATGCCTTG..ACCA.....A.T.TCATTa..TCT.TTA.CTTCAGG...ATGCTTTTCTACT.TTG.AATTGA
Fem1 tgg.ca.t..tc.a.gca.c.aa..atatgccttg..acca.....a.t.tcattha..tct.tta.cttcagg...atgcttttctact.ttg.aattga
Fem1
Fem1 TGG.ca.c..Tc.A.GcA.C.NA..ATATGccTTg..ACCA.....A.T.TcATTa..tct.TTa.CTTCAGG...ATGCTTTTCTACT.TTG.AATTGA
Fem1 ~~~~~NNNNtct.tta.CTTCagg...atgcttttctact.ttg.aaTtGa
Fem1 ~~~~~NNNNaNNNgT.N.TCT.TTA.CTTCAGG...ATGCTTTTCTACT.TTG.AATTGA
Fem1
Fem1
Fem1
Fem1
          2310      2320      2330      2340      2350      2360      2370      2380      2390      2400
gi|75 .....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
Isoti CTGAAGAATTGGTAGCAATGATCTCTTTGTGCGAATCGCTTGTTACAGGTGAATCTTGTTGTTTAGTTTCGAATAATTATGTTTTCTATGATAAATGGAC
Fem1 a.ttgga.c.ctcacaact.a.....aa..ct.ac..ttaaat.t.ct..a...aa.tgataaataa..tcttctga.t.tcc..ctaccaga..t.aatt
Fem1
FEm1
Fem1
FEm1
Fem1 A.TTGA.C.CTCACAACt.A.....AA..CT.ACn.TTAAAT.T.CT..A...AA.TGATAAATAA..TCTTCTGA.T.TCC..CTACCAGA..T.AATT
Fem1 a.ttgga.c.ctcacaact.a.....aa..ct.acg.taaaat.t.ct..a...aa.tgataaataa..TCtc.g.A.tcc..c.CTACC.a.at.T.AATT
Fem1
Fem1 A.TTGc.C.CtCACAACt.A.....AA..CT.a.g.TAAAAt.T.C...A...AA.TGaaaaaNcNNNg.gggggg.T.T.N
Fem1 a.ttgga.c.ctcACAACt.A.....aa..ct.ACG.TTaaat.t.ct..a...AA.tNaaaAATAAa.tCct.cGA.t.t.cc.CTACCAGG..T.AA.t
Fem1 A.TTGA.C.CTCACAACt.A.....AA..CT.ACG.TTAAAT.T.CT..A...AA.TGATAAATAA..TCTTCTGA.T.T.C..CTACCAGA..T.AATT
Fem1
Fem1
Fem1
Fem1
          2410      2420      2430      2440      2450      2460      2470      2480      2490      2500

```

[illegible]

2610 2620 2630 2640 2650 2660 2670 2680 2690 2700

ttgttgcgattgtttt~gctgtgatt~GctttGTTTgttgctGctgtTTTTTGttgcTNttttgtttgtccTAtctgCGGATCTaaTggcgcgGTGTAAT  
TTGTTGcGATTGTTGTTGCTGtTCTtTTGCFTTGTTTTTTTCTAAACTGCTTTTTTCGGAGTTTGCTGTGTTG~TCTGCGGATCTGCTGGCCTTGTG~AT

A horizontal timeline with major tick marks every 10 years, labeled from 2710 to 2800. Minor tick marks are present every 5 years. The timeline is represented by a horizontal line with vertical tick marks.

GGCTTCTGAGCGGcTTGCCttTtAtTgacTTCNNTTGGAtgTgctgttgctgctGcTgcTottgCaagggcttCcCAggAgTTTGttGTcgCTGctgttg  
GGCTTCTGAGCGGTCTGCCTTTGAATTGGCTGCCNNTTGGAAAGGACTTCTGAGGAGTCTGTGTGCTGCTGTTGCCCTGAATACCTCGTGACATGACAC

```

                2810      2820
            ....|....|....|....|....|...
gi|75
Isoti
Fem1
FEm1
Fem1
FEm1
Fem1
Fem1
Fem1
Fem1
Fem1
Fem1 cccNaNaataactCNtNNNNNTTTTTTN
Fem1
Fem1
Fem1
Fem1
Fem1

```

## ANEXO 10.

-Alineamientos para los Isotigs 08283 y 00774/5 en los marcos de traducción correspondientes con las mejores coincidencias por BLASTx en el NCBI. Se incluyen además las secuencias de singletons que también mostraron homología con la secuencia de la proteína DSX así como las secuencias obtenidas a partir de dominios conservados para unión al DNA y de dimerización (DM) para *doublesex*. Se indican en los recuadros las regiones conservadas y a su vez la región variable encargada de la determinación del sexo. Se resalta en amarillo la única porción de secuencia que quedó sin cubrir.

pfam00751 paper2009

pfam00751 consensusDM DNA BD (1-47) Nter

gi|665786878|ref|XP\_008554572.1| PREDICTED: protein doublesex isoform X3 [Microplitis demolitor]



gi|665786873|ref|XP\_008554550.1| PREDICTED: protein doublesex isoform X1 [Microplitis demolitor]

gi|665786875|ref|XP\_008554560.1| PREDICTED: protein doublesex isoform X2 [Microplitis demolitor]

gi|755974785|ref|XP\_011307439.1| PREDICTED: protein doublesex isoform X2 [Fopius arisanus]

gi|755974778|ref|XP\_011307437.1| PREDICTED: protein doublesex isoform X1 [Fopius arisanus]

pfam08828 consensus DSX-dimer dsx dimerization

pfam00751 DMRTA motif dsx Cter consensus

Isotig08283 inv en +2

Isotig00774 en +3

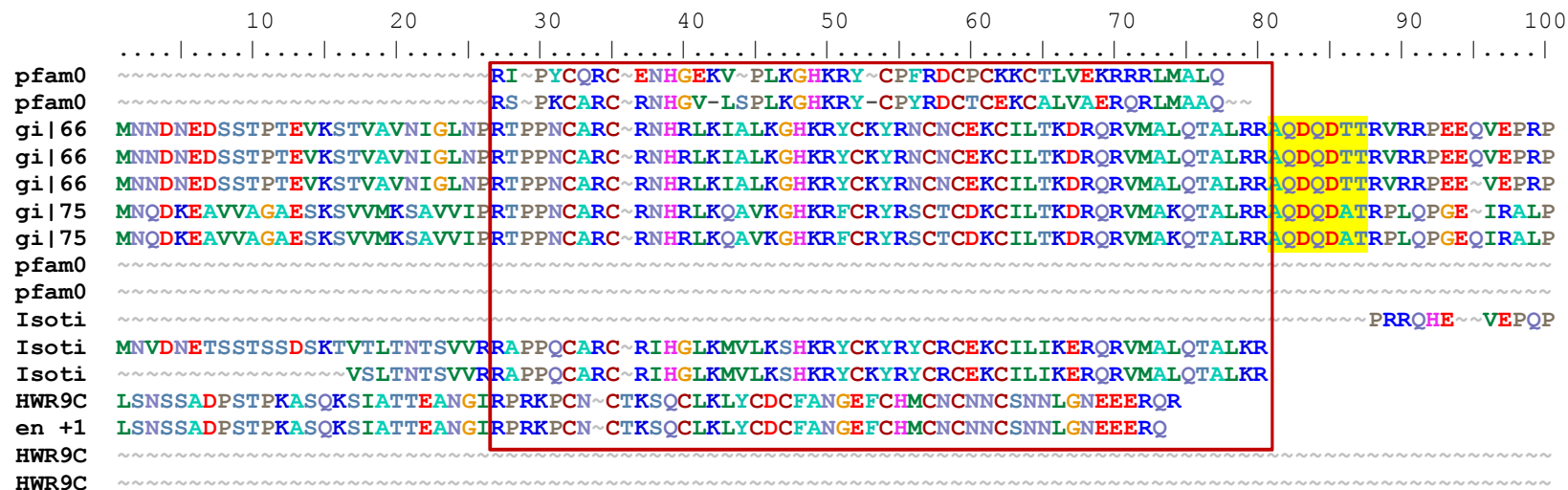
Isotig00775 en +2

HWR9CMJ02JGH1E

en +1 HWR9CMJ02HD6FV

HWR9CMJ02J33 en +1

HWR9CMJ02IRU92 inv en +3



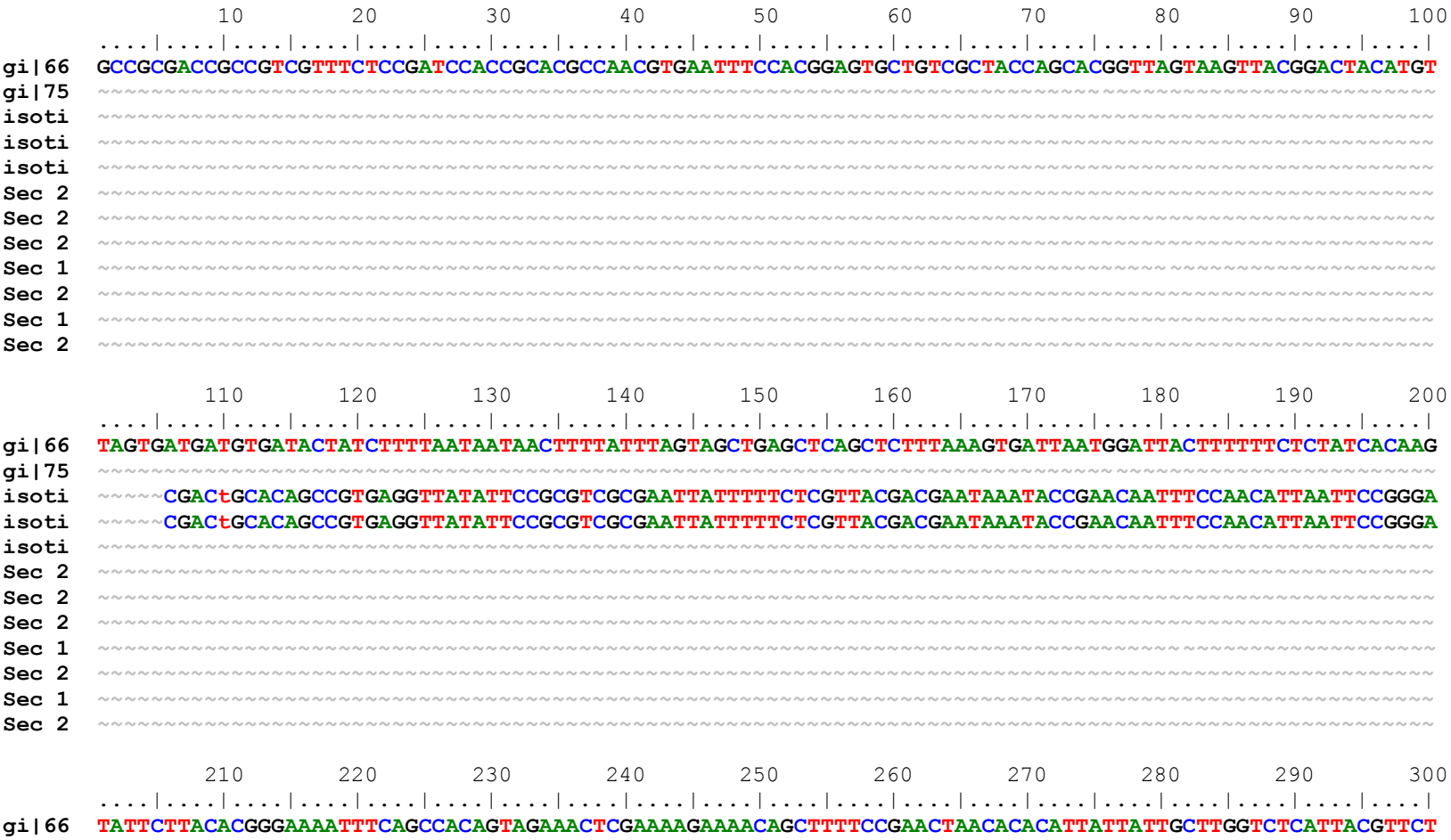
	110	120	130	140	150	160	170	180	190	200
	....	....	....	....	....	....	....	....	....	....
pfam0										
gi 66	LTLDRERLTSASQPTRSLEGSC~DSSSGDSPISNHGSGVH~IVSVPNLQKLSPTLNTHSSSTTQMSEPE									
gi 66	LTLDRERLTSASQPTRSLEGSC~DSSSGDSPISNHGSGVH~IVSVPNLQKLSPTLNTHSSSTTQMSEPE									
gi 66	LTLDRERLTSASQPTRSLEGSC~DSSSGDSPISNHGSGVH~IVSVPNLQKLSPTLNTHSSSTTQMSEPE									
gi 75	IALDGELPISVPQPARSLNSC~DSISADSPFSNQGSTGPHSGIFTIPPSCKLPSPFSSHSTSDTQLSEPRSC									
gi 75	IALDGELPISVPQPARSLNSC~DSISADSPFSNQGSTGPHSGIFTIPPSCKLPSPFSSHSTSDTQLSEPRSC									
pfam0	~LPQ~DVLLDYC~QKLEKFRYPWEM									
pfam0	~SR									
Isoti	MNLNIDRPMSPQPARSLNSCADSGCSSADSPFSNHGGTGSHTGIFTVPSSSKLPQSHNIHSSSATELPE~RNCENVE~VLLEYSVKLEQFWYSWDI									
Isoti										
Isoti										
HWR9C										
en +1										
HWR9C	~SKILNFHFRSTYQ~SPLSTNKNQI~MLQKSGCN~LISRTW~									
HWR9C	~KN~VMDLI~DC~ILGIFSYSWNL									
	210	220	230	240	250	260	270	280	290	300
	....	....	....	....	....	....	....	....	....	....
pfam0										
pfam0										
gi 66	LPLMYVILKDARADIEEATRRIAEVDINEDSLVYFRKMITTVKIFKQIIVKKILNKMWFLLIAV									
gi 66	LPLMYVILKDARADIEEATRRIAEANSEIRAVAF~WKARRMLQDNGGVYNDWYPGANNIVTPTYLGHPYMG									
gi 66	LPLMYVILKDARADIEEATRRIAEANSEIRAVAF~WKARRMLQDNGGVYNDWYPGANNIVTPTYLGHPYMG									
gi 75	LPLMYVILKDAKSDLDEARRRIAEGITEIRSIFV~RKARTMM~GVT~SDVYYNERYTATAGSAG~PTYIGQPPYIGS~VMH~PPL~HLGIPHLVNAHLFTAGVPS									
gi 75	LPLMYVILKDAKSDLDEARRRIAEGITEIRSIFV~RKARTMM~GVT~SDVYYNERYTATAGSAG~PTYIGQPPYIGS~VMH~PPL~HLGIPHLVNAHLFTAGVPS									
pfam0	MPLMYVILKDADADIEEASRRIEEGQHIVNEYSRQHNLN									
pfam0	SPID~VLMKIFPNRKRSLVLELVLRCKGDVLQAIELAL									
Isoti	LPLMYVILKDAKADIDEARRRIAQAVTTY~VLLL~RPQE									
Isoti										
Isoti										
HWR9C										
en +1										
HWR9C	~LYVVFSTDAKQRIEEMSSAI~LLIPPNGKRCVYFVTSLYRSSTQPCPVASCTLHSLALD TYWMTCMHSLISSICIMINS									
HWR9C	LILFFAL~KDGR~DIFYAA~IYETKSTMDLKIIFKRIVDHEOKWPTSILIKIPFINPINPLNFSNOS									

310  
 ....|....|....|.  
 pfam0  
 pfam0  
 gi|66  
 gi|66 PEGSPT  
 gi|66 PEGSPT  
 gi|75 SPPRGPTI  
 gi|75 SPPRGPTI  
 pfam0  
 pfam0  
 Isoti  
 Isoti  
 Isoti  
 HWR9C  
 en +1  
 HWR9C

-Alineamiento para las secuencias nucleotídicas correspondientes a los Isotigs homólogos a *doublesex*. Se incluyen en el alineamiento las secuencias amplificadas por PCR de punto final obtenidas utilizando los cebadores listados (ver materiales y métodos).

gi|665786874|ref|XM\_008556338.1| PREDICTED: Microplitis demolitor protein doublesex (LOC103576202), transcript variant X2, mRNA  
 gi|755974781|ref|XM\_011309136.1| PREDICTED: Fopius arisanus protein doublesex (LOC105269122), transcript variant X2, mRNA  
 isotig00775  
 isotig00774  
 isotig08283 inv  
 Sec 25 dsx  
 Sec 26 dsx  
 Sec 28 dsx  
 Sec 10 dsx  
 Sec 24 dsx

Sec 11 dsx  
Sec 22 dsx



[illegible][illegible][illegible]

Sec 1  
Sec 2  
Sec 1  
Sec 2

510 520 530 540 550 560 570 580 590 600

gi|66 CTGAAAGGTACTCTCAAAGTTCACAGGCGTATCCGGAGGGCTGGACCTTGACTTCCGGGGGCGGTGACTTCCTGTTGTAAC TGCGCGTGTTCCACGTG  
gi|75 ATTAGCACACATAAAAA TACTCCGATACCACACCATTATCCAGTGCCCAAGATGAACCATAGAAAAACACCATATGCACGATCAATTTATCGTAAAT  
isoti TGTGGATGAAGCGAGGAGTTGTAAAAAATTTGATAAAAAAGaGAAATTCAACAAGTGGTCGATGATAAAAAAGCTACCTGGTATCCATGGGTTATGTGTAA  
isoti TGTGGATGAAGCGAGGAGTTGTAAAAAATTTGATAAAAAAGaGAAATTCAACAAGTGGTCGATGATAAAAAAGCTACCTGGTATCCATGGGTTATGTGTAA  
isoti  
Sec 2  
Sec 2  
Sec 1  
Sec 2  
Sec 1  
Sec 2

610 620 630 640 650 660 670 680 690 700

gi|66 ACAGGCGCGAGCGCATACACTTGCTGCGCAACATACACGCGAGTTGAGAAGAAACAAAAAATTTCTAAAAATAATATATATATATATATAATATGATA  
gi|75 TACACACGTACTTAGTAGAGGATAAATACCAGTGGACATGCACGCACCAATTAGAAAAATTATAGAGGAAAAATGAAAAGCCATGTAATACTCTAATGAGA  
isoti GTGGTGACGCGCATGTAAAGTCTGGTAAC TGGGGAGgTTGGACCTTAAC TTCCGGGGGgCGgTGACTTCCTGCTtGGTTTAGCCGTGCGTGTTCCACGT  
isoti GTGGTGACGCGCATGTAAAGTCTGGTAAC TGGGGAGgTTGGACCTTAAC TTCCGGGGGgCGgTGACTTCCTGCTtGGTTTAGCCGTGCGTGTTCCACGT  
isoti  
Sec 2  
Sec 2  
Sec 2  
Sec 1  
Sec 2  
Sec 1  
Sec 2

710 720 730 740 750 760 770 780 790 800

```

gi|66 AGATATTAAGTGCAATTTTATAAAATTGTTCTTGTCAATTTGCGTGGTTCAAACATCACACTTATCAGTGTGGGTTCTGTGTAAGATTCTCTTAGCACTA
gi|75 ACTCGATTCCAATTTCAAATTTTTTAATTAATAAAATTCCTCTCGCTCATTACAAAACCTCTCTGGCAGTTCTCAAGTTTGAAAGTTTTGACGGACAAGTCTG
isoti GACAGGCGCGAGCGCGTTTTCTcACTGCGCCGGATCTGCACGTGGGTATCGGAAGTATTTTGCATCTTTCGGTGGAAATTTGACATCTCGTGGGtGGTTG
isoti GACAGGCGCGAGCGCGTTTTCTcACTGCGCCGGATCTGCACGTGGGTATCGGAAGTATTTTGCATCTTTCGGTGGAAATTTGACATCTCGTGGGtGGTTG
isoti
Sec 2
Sec 2
Sec 2
Sec 1
Sec 2
Sec 1
Sec 2

      810      820      830      840      850      860      870      880      890      900
gi|66 ....|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|
gi|75 AATTAATATCATCAACATATATACGTATATACATATATACACTTACAAATCATTACGTAGCGTGGGAATAAGATAAATAAATAAAGCAGCTGTTGTGATC
isoti CAAGTTCCCCCAAGGTTTGCCGCACTTAAAAATTCACAAGGTTTACATTTATTTGTTTTACTTTATATCACGTTGTGACCTGTTAGTCTACCACGATG
isoti GAGAGAGAAGAAAGGGCCTGACTgtcGAGGatCAACTGGCTTGATTGGTACTGGGGGACTCGGTGGATTGagCACTGAGACGTTGaCTGGGAGCTGGACG
isoti GAGAGAGAAGAAAGGGCCTGACTgtcGAGGatCAACTGGCTTGATTGGTACTGGGGGACTCGGTGGATTGagCACTGAGACGTTGaCTGGGAGCTGGACG
isoti
Sec 2
Sec 2
Sec 2
Sec 1
Sec 2
Sec 1
Sec 2
Sec 2
      910      920      930      940      950      960      970      980      990      1000
gi|66 ....|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|
gi|75 CTCTTCCTTTGGTTGCGGTTTACTATCAACTACAAAGTACCGTAGCAAGAGTATAATCAAGATTTATAGAGCGAATTGAGAAGATCTAATTATCAAAAAC
isoti TGATTGTAATCACGGGATCAATTACTGATGGGGAGAACTCTGAGGACTTGAAGGTATGTAAATAAATAGTCCGTGAATTGATGTGCCAGGGATTGAGTT
isoti TCAGGaCTGCTTTGCGTTGGTGAGGAGgATGGGGAAACCGGGcGTTCTCACcGGGGACGTGGTCTGAGGGCAGTCGGGATTTTTTTGAGTGGCAGAATAAA
isoti TCAGGaCTGCTTTGCGTTGGTGAGGAGgATGGGGAAACCGGGcGTTCTCACcGGGGACGTGGTCTGAGGGCAGTCGGGATTTTTTTGAGTGGCAGAATAAA
isoti
Sec 2
Sec 2

```

```

Sec 2
Sec 1
Sec 2
Sec 1
Sec 2
Sec 2
TCAGGNCTGCTTTGCGTTGGTGAGGAGGATGGGGAACCGGGGAGTTCTACCCGGGGACGTGGTCTGAGGGCAGTCGGGATTTTTTGTAGTGGCAGAATAAA

      1010      1020      1030      1040      1050      1060      1070      1080      1090      1100
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
gi|66 ATGAAAATCATCCTTAGAGAGCGAAATGTATGACCAATAATTACGGAGTGTCAAATGGATGTGATAGGAGGTGCTAGTGATTGTTAGAACAGACTCGTCA
gi|75 GCTGAGAGCAACGCGAAAAGCGAGATCGAAGGAGAAAGTGCGTGAGGCGCGGAGTGAAACAATTAATCAAGTACATTGAATTATCAGTGGTGCTATTCAAT
isoti CATCAGCGAACGTCGAAGTGCTCAAGTGATGAATGAAGTACAATTACCAGTGATACTAATCGTGATATCCAAAGtCcGATGAAATTACAAATTATAGTAA
isoti CATCAGCGAACGTCGAAGTGCTCAAGTGATGAATGAAGTACAATTACCAGTGATACTAATCGTGATATCCAAAGtCcGATGAAATTACAAATTATAGTAA
isoti
Sec 2
Sec 2
Sec 2
Sec 1 CGACAGCGAACGCTGCAGTGCAGCAAGTGATTGATGCCGTACAATTACCAGTGATACTAATCGTGATATCCAAAGTCCGATGAAATTAC~TTATAGAGA
Sec 2 CATCAGCGAACGTCGAAGTGCTCAAGTGATGAATGAAGTACAATTACCAGTGATACTAATCGTGATATCCAAAGNCCGATGAAATTACAAATTATAGNAA
Sec 1 CATCANCGAACGTCNAAGTGCTCAAGTGATGAATGAAGTNCNNTTACNNNTGATACTANTCGTGATNTCNNNAGNCCGATGAAATTACAAATTATAGTAN
Sec 2 CATCAGCGAACGTCGAAGTGCTCAAGTGATGAATGAAGTACAATTACCAGTGATACTAATCGTGATATCCAAAGNCCGATGAAATTACAAATTATAGTAA

      1110      1120      1130      1140      1150      1160      1170      1180      1190      1200
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
gi|66 GTAAATAAACAATTAAATGATTTTAAAGTGTTGAACATTAATATGAATAACGATAACGAAGATTCAACAACCGACTGAAGTTAAGTCAACGGTAGCTG
gi|75 AAAC TAGTCTCGTTAATCAGTTGGAGTGTTAAAAAATTAATAATGAATCAAGATAAGGAGGCAGTAGTGGCTGGTGCTGAGAGTAAATCGGTTGTTATGA
isoti ATTCTAAACAACAGTGATCAGAACATTGACAAAATGAATCTCGTtGAACtTGTCTCG~TTAACAAATACAA
isoti ATTCTAAACAACAGTGATCAGAACATTGACAAAATGAATGTGGATAACGAGACAAGTTCAACGAGTTCGGACTCTAAAACGGTCACcTTAACAAATACAA
isoti
Sec 2
Sec 2
Sec 2
Sec 1 ATTcacaACAACAGTGATCAGaacaTGACAACgtgAaTgTGATAACGaGACAAGTTCAACgNgTTCGgAcccTaAaaCGGtcgccTTatCAAATACaa
Sec 2 ATTCTAAACAACAGTGATCAGAACATTGACAAAATGAATGTGGATAACGAGACAAGTTCAACGAGTTCGGACTCTAAAACGGTCACCTTAACAAATACAA
Sec 1 ATTCTAAACAACAGTGATCANAAACATTGACAAAATGAATGTGGATAACNAGACNAGTTCAACGAGTTCGGACTCTAAAACGGTCACCTTAACAAATACAA
Sec 2 ATTCTAAACAACAGTGATCAGAACATTGACAAAATGAATGTGGATAACGAGACAAGTTCAACGAGTTCGGACTCTAAAACGGTCACCTTAACAAATACAA

      1210      1220      1230      1240      1250      1260      1270      1280      1290      1300

```



```

...|. ...|. ...|. ...|. ...|. ...|. ...|. ...|. ...|. ...|. ...|. ...|. ...|. ...|.
gi|66 TAAATATAGGGTTAAATCCAAGAACTCCGCCAAATTGTGCACGGTGTCTGTAACCATCGTTAAAAATAGCTCTCAAGGGCCATAAAAGATACCTGCAAGTA
gi|75 AAAAGTGCTGTGGTGATACCACTGACACCACCAAAATTGTGCGCGTTGTCTGTAATCATCGGTTGAAACAAGCCGTCAAGGGACACAAGAGATTTTGCCGATA
isoti GCGTTGTACGTCGTGCACCAACCACAATGTGCACGCTGCCGTATTTCATGGATTGAAAATGGTCCTCAAGTCCCACAAAAGATATTGCAAATACAGATACTG
isoti GCGTTGTACGTCGTGCACCAACCACAATGTGCACGCTGCCGTATTTCATGGATTGAAAATGGTCCTCAAGTCCCACAAAAGATATTGCAAATACAGATACTG
isoti
Sec 2
Sec 2
Sec 2
Sec 1 GcgTtgTccgTCgTGcAccaCCgcAatgTGcAcGctgCcGTAttcAtggatTgaaaAtGgtCctcatgtccgaCaaaagataattgcaaatacagatactg
Sec 2 GCGTTGTACGTCGTGCACCAACCACAATGTGCACGCTGCCGTATTTCATGGATTGAAAATGGTCCTCAAGTCCCACAAAAGATATTGCAAATACAGATACTG
Sec 1 GCGTTGTACGTCGTGCACCAACCACAATGTGCACGCTGCCGTATTTCATGGATTGAAAATGGTCCTCAAGTCCCACAAAAGATATTGCAAATACAGATACTG
Sec 2 GCGTTGTACGTCGTGCACCAACCACAATGTGCACGCTGCCGTATTTCATGGATTGAAAATGGTCCTCAAGTCCCACAAAAGATATTGCAAATACAGATACTG

1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
...|. ...|. ...|. ...|. ...|. ...|. ...|. ...|. ...|. ...|. ...|. ...|. ...|. ...|.
gi|66 TCGAAACTGTAATTGTGAGAAGTGTAATCTTACTAAAGATCGTCAAAAGAGTTATGGCCTTACAAACTGCGCTCAGGAGAGCTCAGGATCAAGATACGACA
gi|75 CCGATCGTGTAATGTGACAAGTGTAATCCTTACCAAGGATCGACAGAGGGTGATGGCGAAACAGACAGCCCTAAGGAGGGGCACAGGATCAGGATGCAAC~
isoti CAGATGTGAGAAGTGTAATCTTATCAAaGAGCGACAGAGGGTGATGGCATTGCAAAACGGCGTTGAAGAGAGC
isoti CAGATGTGAGAAGTGTAATCTTATCAAaGAGCGACAGAGGGTGATGGCATTGCAAAACGGCGTTGAAGAGAGC
isoti
Sec 2
Sec 2
Sec 2
Sec 1 cagcTgttgagaagTGtattcTtAtacacAGgCgACagaggggTgAtggcaTtgcacACggCgttggaAGAGaGCAAGGAgaGCGCAGGAtCAGGATGCAACa
Sec 2 CAGATGTGAGAAGTGTAATCTTATCAAAGAGCGACAGAGGGTGATGGCATTGCAAAACGGCGTTGAAGAGAGCACAGGATC~AGGATCAGGATGCAACA
Sec 1 CAGATGTGAGAAGTGTAATCTTATCAAAGAGCGACAGAGGGTGATGGCATTGCAAAACGGC
Sec 2 CAGATGTGAGAAGTGTAATCTTATCAAAGAGCGACAGAGGGTGATGGCATTG

1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
...|. ...|. ...|. ...|. ...|. ...|. ...|. ...|. ...|. ...|. ...|. ...|. ...|. ...|.
gi|66 AGAGTTCGAAG~GCC~AGAGGAGGTGGAACCACGACCATTAACCTTTAGACAGAGAAAGGCTAACG~TCGGCATCCCAACCAACGCGCAGTTTGGA
gi|75 ~GAGGCCGCTGCAGCCTGGAGAGCAGATCAGAGCACTACCCATAGCCCTGGATGGGGA~GCTTCCGATATCCGTGCCACAGCCTGCCCGTAGCCTAGA
isoti
isoti
isoti
isoti
isoti
Sec 2

```

```

Sec 2 ~~~~~TGCCACGAAGACAACATGAGGTGGAACCACAACCACAGAATCTGAATATTGAATCC~CCGATGTCAGTACCACAGCCAGCTCGTAGCCTAGA
Sec 2 ~~~~~
Sec 1 aGAGcaCGttCCAgacGagaacgagGGgGTGaaccAcaaccaatgaatctGAACATTTAcAgAg~CCGaTGgCcGtACCaccGCCggCTCGtAGTCTaga
Sec 2 AGAGCACGTTCCAGGCGAGAACGAGGGGGTGAACCACAACCAATGAATCTGAACATTGACCACN~CCGATGGCCGTACCACC GCCGGCTCGTAGTCTAGA
Sec 1
Sec 2

          1510      1520      1530      1540      1550      1560      1570      1580      1590      1600
gi|66  ....|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|
gi|75  AAA TAGCTGTGA ~~~~~ CAGCAGTAGTGGAGATTCAACGATAAGTAATCACGGTGGTTCGGGTGTACA ~~~~~ CATCGTTAGTGTTC ~~~ CCA
isoti
isoti
isoti gaatagctgtgctgattctggatgcagcagcagtgccgattcgccattcagcaatcacggcgggcacaggttcacacactgggtatatttcacagttccatca
Sec 2 ~~~~~ AGCTGTGCTGATTCTCGGATGCAGCANCAGTCCGATTCTGCCATTNANCAATCANGGCGGCACAGGTTACACACTGGTATTTTCACAGTTCCATCA
Sec 2 GAA TAGCTGTGCTGATTCTGGATGCAGCAGTAGTCCGATTCTCGGATTTCAGCAATCACGGAGGCACAGGTTGACACACTGGAATT ~~~~~
Sec 2 ~~~~~
Sec 1 GAA TAGCTGTGCTGATTCTcgATGCAGCAGcAGTCCGATTCTGACANN CAGCAATCaCGGCGGCACAGgtTTCACACACTGGTaTTTTTCACtTtTCCATCA
Sec 2 GAA TAGCTGTGCTGATTCTCGA ~~~~~ AGCAGCC ~ATT ~~~~~ CANCAATCANGGCGGCNCAGGTTTCACACACTGGTATTTTCACNNTTCCATCA
Sec 1
Sec 2

          1610      1620      1630      1640      1650      1660      1670      1680      1690      1700
gi|66  ....|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|...|
gi|75  ACC TACAGAAACTATCACCAACTTTAAACACTCATTTCATCGTCAACAACGCAAA TGTCCGAACCCGAGAGTTGTGA ~~~ AAATGTTGAAGTACTATTGGA
isoti TCCTGC ~~~ AAAC TACCACCGTCCCTTCAGCAGTCACTCGACTTCGGATACACAATTATCTGAGCCGAGGAGTTGTGAGTCAAGTGATGACATGTTGCTGGA
isoti
isoti tcaagc ~~~ aaactaccgcagtccccacaacatccactcgtcgtcgggtacagaattgccagaac ~~~ gaaattgtgaaa ~acgtg ~~~ gaagtcoctgttgga
Sec 2 TCAAGC ~~~ AAAC TACCGCANTCCCACNACATCCACTCGTCNTCNGCTACAGAA TTGCCAGAAC ~~~ GAAATTGTGAAA ~ACGTG ~~~ NAANTCCTGTTGGA
Sec 2 ~~~~~
Sec 2 ~~~~~ AATTGCCAGAAC ~~~ GAAATTGTGAAA ~ACGTG ~~~ AAAGTA ~~~ GTTGGA
Sec 1 TCAAGC ~~~ AAac TACCGCggTCCCaaaaCaccgAcTCGTCGTCgGcTAcAAaATTGCCaGAAC ~~~ a ttttGAAACACgtgATGaaGTCCTgTTGGg
Sec 2 TCAAGC ~~~ AAAC TACC
Sec 1
Sec 2

```

```

1710      1720      1730      1740      1750      1760      1770      1780      1790      1800
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
gi|66      ATGTAGTACAAAACCTCTCGAACGATTTTGGTATTCTGGGATATGTTGCCCTTAATGTATGTAATATTAAAAGATGCAAGAGCAGACATTGAAGAAGCA
gi|75      GTACAGTGTCTAAATTGCTCGACCAATTTTGGTATACGTCGGAAATATTACCGCTTATGTACGTCATACTAAAAGACGCTAAGTCAGATCTCGACGAAGCG
isoti
isoti
isoti      gtacagtgtaagttactcgagcaatTTTTGGTATTcatgggacatattaccgctgatgtacgtcatactcaaggatgccaaaggcagatatcgacgaagct
Sec 2      GTACAGTGTCAAGTTACTCGAGCAATTTTGGTATTcatGGGACATATTACCGCTGATGTACGTCATACTCAAGGATGCCAAGGCAGATATCGACGAAGCT
Sec 2      GTACAGTGTCAAGTTACT~GAGCAATTTTGGTATTcatGGGACATATTACCGCTGATGTACGTCATACTCAAGGATGCCAAGGCAGATATCGACGAAGCT
Sec 2      GTAC~GTGTCAGTTACTCGAGCAATTTTGGTATTcatGGGACATATTACCGCTGATGTACGTCATACTCAAGGATGCCAAGGCAGAAcACTAGACTCCT
Sec 1      GaAcaggAtNNNNNNNNNN
Sec 2
Sec 1
Sec 2

1810      1820      1830      1840      1850      1860      1870      1880      1890      1900
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
gi|66      ACAAGACGTATAGCAGAAAGCTAACAGTGAATTCGAGCCGTGGCGTTTTGGAAAGCGAGACGAATGCTGCAAGACAACGGAGAGTTTACTACAATGATT
gi|75      AGGAGACGCATAGCCGAAGGCATTACTGAGATACGTTCCATATTTGTGAGGAAGGCCGAGAACAATGATGGGTGTTACTAGTGATGTTTACTACAATGAAC
isoti
isoti
isoti      cgaagacgcattgcacaagca~gtaacgacatacgtgctgttgctgtgatgaaggccacaagaatgatgac
Sec 2      CGAAGACGCATTGCACAAGCTAGTAACGACATACGTGCTGTTGCTgtgatGA~GGCCAA
Sec 2      CGAAGACGCATTGCACAAGCTAGTAACGACTTACGTGCTGTTGCTGTGATGAAGGCCACAAGAATATANAACCTTCNTTTTGCCAAAAACATCTTTGTTA
Sec 2      GGAAaaaaaCaTcCACAAG~TGTGAAA~CCACAACCGTCGTGACAcacaccaagagtggaggagaCaagNN
Sec 1
Sec 2
Sec 1
Sec 2

1910      1920      1930      1940      1950      1960      1970      1980      1990      2000
....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
gi|66      GGTACCCCTGGCGCCAATAATATTGTGACGCCAACGTATCTTGGTCACCCACCATACATGGGAGGAGGAATAGCACCATCGCATACATCTGTACATGTTGG
gi|75      GGTACACCGCAACAGCTGGGAAGTGCTGGACCCACCTACATTGGTCAACCACCGTACATTGGCAGTGTTCATGCATCCTCCTCTGCACTTAGGAATTCCTCA
isoti
isoti
isoti

```

Sec 2  
Sec 2 ATGATTCCGTGCCATCAAGTGTCTGCTCACAATTTTGATATTTTCACTGCTGAATTCTCAGTGCCACATTTGTTCCCAATATTTGTGTANAGTAATTT  
Sec 2  
Sec 1  
Sec 2  
Sec 1  
Sec 2

2010 2020 2030 2040 2050 2060 2070 2080 2090 2100  
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|  
gi|66 ACCTCTTCCCCATTTATTAAATGCTCAAGTGTCTGCGAGTAGGATACCCTCGAGTCCGGACAGCCCTCCTGAGGGGTCACCTACGTAGTTCAGATTATCC  
gi|75 CTTAGTAAATGCCCATCTATTCACTGCTGGAGTACCATCGAGTCCACCCAGAGGACCCACCATCTAGTCCAGATTACCGTGAAGTCAATCAAAAAAGCTC  
isoti  
isoti  
isoti  
Sec 2  
Sec 2 CNCATTATTTTCTTCGTGGANNNTACNATGANGCNNNNATGAGGAGTGACNTCAATGAAGATGATCGATCANTGGAGATGATGCAACGCCTGCNAGATG  
Sec 2  
Sec 1  
Sec 2  
Sec 1  
Sec 2

2110 2120 2130 2140 2150 2160 2170 2180 2190 2200  
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|  
gi|66 TGAACCAATAGAAAGTGCTCAATAATAACAGTCACAGGTTGAGTCTAATTCAAATTATTTAACATCTGCGTAACCAATTGAACCAATCAACACATCTAGC  
gi|75 GTAAGAGAGGTCAGAGGCTGAGTTCTACGGAAGGAAAAATAAAATACGGTGACTTGAGCAATCAAAAGAGGTGCATAATGAATAGTTGGAGGTTTT  
isoti  
isoti  
isoti  
Sec 2  
Sec 2 ATGACTGCTCANAAGATTTTCTCCTCGNTGTTTTTCGGGACAACCTTGTTCTAGTCTATTGACGTTTATTTGATAGTATCTATCGGTTGTAGTTGAAAAT  
Sec 2  
Sec 1  
Sec 2  
Sec 1  
Sec 2

```

                2210      2220      2230      2240      2250      2260      2270      2280      2290      2300
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
gi|66  TGTATAATTTTGGTGAATAGTGTCTGAGAAATTGAACGAATTTTGTATATTTTAGTTATTTATTTAAATTTGTTATACCTTGCGTCGTAGAGGTAACA
gi|75  TTCTTTTATTATTATTCCAATCGCAAAATCGTTTGCACCTCCACCAGTACAGAGGAATAGTAGAAGGAAAAAAAAAATGAGGTTGGACACACGTGTTACGA
isoti
isoti
isoti
Sec 2
Sec 2  TATTGCGANTCATGAACGTTCCACTATGAANATTTCACTGTACCAGTGGCCTAAGTCGGGGTAAATGTTAAGAAGATTGGGGAGGACAATAAGCTAGTAN
Sec 2
Sec 1
Sec 2
Sec 1
Sec 1
Sec 2

                2310      2320      2330      2340      2350      2360      2370      2380      2390      2400
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
gi|66  AAAAGTTTTTAACTGAAATATGATTAAACATTAAACGATAATGAAATATAAAAATTTTGTATATAAAGTAGTCAAAATGTAGATAATGAGAATGA
gi|75  CTTTCGGTGTAATTAATGCAATAATAAAATGGTAAATAGTGGAATTTTTTATCAACAATGGGACATTTCCAAAGTCACAAACCTATCTTCTCTCTTT
isoti
isoti
isoti
Sec 2
Sec 2  CGANATACNTGCNG
Sec 2
Sec 1
Sec 2
Sec 1
Sec 1
Sec 2

                2410      2420      2430      2440      2450      2460      2470      2480      2490      2500
      ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
gi|66  AATATTTATAATTTTATTATCGCTAATTAAGAGACAAATAAGGAAAAAAGTCATTTGATAGTGAGAAGAACTGTAAGGGGACCAATATCTTCAAGAACT
gi|75  TTTTATCATTTTTTCTCTTCCGTTGTGATATTCAGAGGAGAATATTCATTCAAAGTACTGTATCTAGTTAACCATTCTTATTTGATCATCATCGGCCA
isoti
isoti
```

isoti  
Sec 2  
Sec 2  
Sec 2  
Sec 1  
Sec 2  
Sec 1  
Sec 2

```
          2510      2520      2530      2540      2550      2560      2570      2580      2590      2600
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
gi|66  AGTGA AAAATTAA GTTAAAAA TTTATATTATAAATATTTGTAACAGCTTC AAAATTTCCGT CACAGGTAA TAGTTTAAAAA CTATGATTTTATTC CCCTTC
gi|75  TCATCGAATGATGATTATTGATAACGTGGTGTATTAAACATGATTATAAGTTATGTCACAGTTTTGACGTATGTACAGATATATGAGTTGGTTTATTGAA
```

isoti  
isoti  
Sec 2  
Sec 2  
Sec 2  
Sec 1  
Sec 2  
Sec 1  
Sec 2

```
          2610      2620      2630      2640      2650      2660      2670      2680      2690      2700
.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|
gi|66  ATCA TTAAGATATACTCTATCTTATTTTTTTTTTATTGCAA
gi|75  AAAA TTTCTCCCAACATTTTTTCA TTGACAGAGTTTAGGAAAA TGTT CAGTTACTCCCC TTTATATTTTCACTGTTATTGTTTTTTC AAGGGACAACATTT
```

isoti  
isoti  
isoti  
Sec 2  
Sec 2  
Sec 2  
Sec 1  
Sec 2  
Sec 1

gi | 66

isoti

isoti

isoti

Sec 2

**Sec 2**

## Sec 2

**Sec 1**

**Sec 2**

Sec 1

## Sec 2

gi | 66

gi | 75

isoti

isoti

isoti

**Sec 2**

## Sec 2

## Sec 2

**Sec 1**

**Sec 2**

Sec 1

Sec 2

gi | 66

gi | 75

isoti

```
gi|66
gi|75
isoti
isoti
isoti
Sec 2
Sec 2
Sec 2
Sec 1
Sec 2
Sec 1
Sec 2
```

TTGTTTCCGTAGCTGAATAACGATAAGTGTCAGCTCCACTAGATCGTGAAATGTTGGCATTATGCACATTAGTTATAAATTTCGAAATGTGTAGCACTT

3110 3120 3130 3140 3150 3160 3170 3180 3190 3200

TGCAACGTGTCA TGGTCGTCGACGCTGTAAAAATCCGAGGACAATTGATTATCAC TTTTTCTCCAATAA TTTCTTCTATTTAGTAAC TTGTAGTTAGTTA



**Sec 2**

gi | 75

isoti

isoti

isoti

Sec 2

Sec 2

Sec 2

Sec 1

Sec 2

Sec 1

Sec 2

3210 3220 3230 3240 3250 3260 3270 3280 3290 3300

GGATATGAAGTATAAAATGCGGTCTATCAATATGAGTATAAAAAGAAATAATAATGAGGCAGGATTGTTGCATGATTTTTTGTTCATTCCTAACAGGGA

gi | 66

gi | 75

isoti

isoti

isoti

Sec 2

Sec 2

Sec 2

Sec 1

Sec 1

Sec 2

Sec 2

Sec 1

Sec 1  
Sec 2

SEC 2

3310 3320 3330 3340 3350 3360 3370 3380 3390 3400

AGTTGCATTCTACTTTAGCAATTGTTCCGAAGTTTTTTCATACACCTCCAGACAGAATAAAATAATGCAAGTTCCTTGTGACGAATAAAAAAAAAATGTCTG

gi | 66

gi | 75

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

3410      3420      3430      3440      3450      3460      3470      3480      3490      3500

TGGACTTATTTTATCTCTACCTTGTAAATCTTTTATGCAGACTATTATAGTTATTTATTGATAAAGACTTCAGTGGACGATTGTGAGACATCAATCTGAA

```
gi|66
gi|75
isoti
isoti
isoti
Sec 2
Sec 2
Sec 2
Sec 1
Sec 2
Sec 1
Sec 2
```

GAAA TTGT CAAAAAG TCCAGTTAG AATTAAATAAGAAAA CCATGTTGA ACCCATAA TCCTTAAACGCTGTAGGGCCATAAAGCTTGGAA TCAAAA

```
gi|66
gi|75
isoti
isoti
isoti
Sec 2
Sec 2
Sec 2
Sec 1
```

AA TAAAGTTAAAAAAAGGCAAAACGTTTTTGTGAATTTCAATGTATTCAATGCATTAAATCATCATAACTGTAAATATAGTATGCGATGTAA TAAAA

Sec 2  
Sec 1  
Sec 2

                  3710          3720  
      ....|....|....|....|. .  
gi|66  
gi|75 ACAA**TTTC**GT**TTCT**AT**TTCAA**  
isoti  
isoti  
isoti  
Sec 2  
Sec 2  
Sec 2  
Sec 1  
Sec 2  
Sec 1  
Sec 2

## ANEXO 11.

- Análisis estadístico para perfiles de expresión a lo largo del desarrollo (Statistica®)
- *Feminizer*

Variante 1: Isotigs 06880/06880b

**ANOVA de 2 factores:** Estadio del desarrollo y condición (virgen: muestra compuesta por machos; control: muestra compuesta por machos y hembras).

Efecto	Univariate Tests of Significance for NRQ Variante 1 Sigma-restricted parameterization Effective hypothesis decomposition				
	SS	Grados de Libertad	MS	F	p
Intercepto	8452,86	1	8452,859	627,4985	0
Estadio	26646,61	4	6661,653	494,5282	0
Condición	4879,65	1	4879,655	362,2415	0
Estadio*Condición	16881,02	4	4220,256	313,291	0
Error	269,41	20	13,471		

Las interacciones entre factores son significativas, por lo que se procede a analizar los factores por separado.

**ANOVA de un factor:** Estadio del desarrollo

Efecto	Univariate Tests of Significance for NRQ Sigma-restricted parameterization; Effective hypothesis decomposition				
	SS	Grados de Libertad	MS	F	p
Intercepto	8452,86	1	8452,859	9,5924	0,004775
Estadio	26646,61	4	6661,653	7,559719	0,00039
Error	22030,09	25	881,204		

La variación observada entre muestras puede atribuirse fundamentalmente al estadio del desarrollo. Para poder conocer qué estadios son los responsables de estas diferencias se llevó a cabo el Test de Bonferroni.

Bonferroni test; variable NRQ Variante 1

Probabilities for *Post Hoc* Tests

Error: Between MS = 881,20, df = 25,000

Estadio	Huevo 76,3	Larva 1 4,3113	Larva 3 0,03783	Pupa 0,1855	Adulto 3,094
Huevo		0,002957	0,001551	0,001586	0,002461
Larva 1	0,002957		1	1	1
Larva 3	0,001551	1		1	1
Pupa	0,001586	1	1		1
Adulto	0,002461	1	1	1	

**ANOVA de un factor:** Condición

- Análisis conjunto de estadios

Efecto	Univariate Tests of Significance for NRQ Variante 1 Sigma-restricted parameterization Effective hypothesis decomposition				
	SS	Grados de Libertad	MS	F	p
Intercept	8452,86	1	8452,859	5,404018	0,027565
Condición	4879,65	1	4879,655	3,119624	0,088257
Error	43797,05	28	1564,18		

Al analizar en conjunto el resultado es marginalmente significativo. Se procede a analizar para cada estadio la diferencia en el factor condición

- Huevos

Efecto	Univariate Tests of Significance for Var3 (Spreadsheet71) Sigma-restricted parameterization Effective hypothesis decomposition				
	SS	Grados de Libertad	MS	F	p
Intercept	34930,29	1	34930,29	530,9913	0,000021
Condición	21703,2	1	21703,2	329,9203	0,000054
Error	263,13	4	65,78		

- Larva 1

Effect	Univariate Tests of Significance for NRQ Variante 1 Sigma-restricted parameterization Effective hypothesis decomposition				
	SS	Grados de Libertad	MS	F	p
Intercept	111,5256	1	111,5256	129,6365	0,000339

Condición	54,9885	1	54,9885	63,9182	0,001327
Error	3,4412	4	0,8603		

- Larva 3

Efecto	Univariate Tests of Significance for NRQ Variante 1 Sigma-restricted parameterization Effective hypothesis decomposition				
	SS	Grados de Libertad	MS	F	p
Intercepto	0,008588	1	0,008588	158,5508	0,000229
Condición	0,006468	1	0,006468	119,4123	0,000398
Error	0,000217	4	0,000054		

- Pupa

Efecto	Univariate Tests of Significance for NRQ Variante 1 Sigma-restricted parameterization Effective hypothesis decomposition				
	SS	Grados de Libertad	MS	F	p
Intercepto	2600,46	1	2600,46	550,7228	0,00002
Condición	205,253	1	205,253	43,4683	0,002741
Error	18,888	4	4,722		

- Adulto

Efecto	Univariate Tests of Significance for NRQ Variante 1 Sigma-restricted parameterization Effective hypothesis decomposition				
	SS	Grados de Libertad	MS	F	p
Intercepto	0,294817	1	0,294817	166,3125	0,000208
Condición	0,032561	1	0,032561	18,3682	0,012789
Error	0,007091	4	0,001773		

Variante 2: Isotigs 07202/06260

**ANOVA de 2 factores:** Estadio del desarrollo y condición (virgen: muestra compuesta por machos; control: muestra compuesta por machos y hembras).

Efecto	Univariate Tests of Significance for NRQ fem1 Sigma-restricted parameterization
--------	--

	Effective hypothesis decomposition				
	SS	Grados de Libertad	SC (MS)	F	P
Intercepto	815,251	1	815,2506	499,0307	0,000000
Estadio	1904,659	4	476,1648	291,4697	0,000000
Condición	27,681	1	27,6806	16,9439	0,000536
Estadio*Condición	188,032	4	47,0081	28,7746	0,000000
Error	32,673	20	1,6337		

Las interacciones entre factores son significativas, por lo que se procede a analizar los factores por separado.

#### ANOVA de un factor: Estadio del desarrollo

Effect	Univariate Tests of Significance for NRQ (anova 2 factores fem 1) Sigma-restricted parameterization Effective hypothesis decomposition				
	SS	Degr. of Freedom	MS	F	p
Intercept	815,251	1	815,2506	82,05466	0
Estadio	1904,659	4	476,1648	47,9258	0
Error	248,386	25	9,9355		

La variación observada entre muestras puede atribuirse fundamentalmente al estadio del desarrollo. Para poder conocer qué estadios son los responsables de estas diferencias se llevó a cabo el Test de Bonferroni.

Bonferroni test; variable NRQ (anova 2 factores fem 1) Probabilities for Post Hoc Tests Error: Between MSE = 9,9355, df = 25,000					
Estadio	Huevo 0,32833	Larvas 1 0,25967	Larvas 3 4,4367	Pupa 20,819	5 0,22167
Huevo		1	0,329598	0	1
Larva 1	1		0,30384	0	1
Larva 3	0,329598	0,30384		0	0,290387
Pupa	0	0	0		0
Adulto	1	1	0,290387	0	

#### ANOVA de un factor: Condición

Effect	Univariate Tests of Significance for NRQ (anova 2 factores fem 1) Sigma-restricted parameterization				
	SS	Degr. of Freedom	MS	F	p

	Effective hypothesis decomposition				
	SS	Degr. of Freedom	MS	F	p
Intercept	815,251	1	815,2506	10,74028	0,002798
Condicion	27,681	1	27,6806	0,36467	0,550785
Error	2125,365	28	75,9059		

En este caso al analizar todos los estadios en conjunto el resultado es marginalmente significativo. Se procede a analizar para cada estadio la diferencia en el factor condición

- Huevos

Effect	Univariate Tests of Significance for NRQ (Spreadsheet37) Sigma-restricted parameterization Effective hypothesis decomposition				
	SS	Degr. of Freedom	MS	F	p
Intercept	0,646817	1	0,646817	317,2744	0,000058
Condicion	0,008513	1	0,008513	4,1756	0,110512
Error	0,008155	4	0,002039		

- Larvas 1

Effect	Univariate Tests of Significance for NRQ (Spreadsheet37) Sigma-restricted parameterization Effective hypothesis decomposition				
	SS	Degr. of Freedom	MS	F	p
Intercept	0,404561	1	0,404561	106,1932	0,0005
Condicion	0,138017	1	0,138017	36,228	0,003838
Error	0,015239	4	0,00381		

- Larvas 3

Effect	Univariate Tests of Significance for NRQ (Spreadsheet37) Sigma-restricted parameterization Effective hypothesis decomposition				
	SS	Degr. of Freedom	MS	F	p
Intercept	118,1041	1	118,1041	34,34438	0,004232
Condicion	10,2809	1	10,2809	2,98966	0,158857
Error	13,7553	4	3,4388		



## - Pupa

Effect	Univariate Tests of Significance for NRQ (Spreadsheet37) Sigma-restricted parameterization Effective hypothesis decomposition				
	SS	Degr. of Freedom	MS	F	p
Intercept	2600,46	1	2600,46	550,7228	0,00002
Condicion	205,253	1	205,253	43,4683	0,002741
Error	18,888	4	4,722		

## - Adulto

Effect	Univariate Tests of Significance for NRQ (Spreadsheet37) Sigma-restricted parameterization Effective hypothesis decomposition				
	SS	Degr. of Freedom	MS	F	p
Intercept	0,294817	1	0,294817	166,3125	0,000208
Condicion	0,032561	1	0,032561	18,3682	0,012789
Error	0,007091	4	0,001773		

## - Doublesex

**ANOVA de 2 factores:** Estadio del desarrollo y condición (virgen: muestra compuesta por machos; control: muestra compuesta por machos y hembras).

Effect	Univariate Tests of Significance for dsx NRQ (Spreadsheet2) Sigma-restricted parameterization Effective hypothesis decomposition				
	SS	Degr. of Freedom	MS	F	p
Intercept	5124,93	1	5124,93	923,5713	0
Estadio	12753,62	4	3188,405	574,5872	0
Condición	539,61	1	539,608	97,2435	0
Estadio*Condición	2262,52	4	565,63	101,933	0
Error	110,98	20	5,549		

Las interacciones entre factores son significativas, por lo que se procede a analizar los factores por separado.

**ANOVA de un factor:** Estadio del desarrollo

Effect	Univariate Tests of Significance for dsx NRQ (Spreadsheet2) Sigma-restricted parameterization Effective hypothesis decomposition				
	SS	Degr. of Freedom	MS	F	p
Intercept	5124,93	1	5124,93	43,98163	0,000001
Estadio	12753,62	4	3188,405	27,36257	0
Error	2913,11	25	116,524		

La variación observada entre muestras puede atribuirse fundamentalmente al estadio del desarrollo. Para poder conocer qué estadios son los responsables de estas diferencias se llevó a cabo el Test de Bonferroni.

Bonferroni test; variable dsx NRQ (Spreadsheet2) Probabilities for Post Hoc Tests Error: Between MS = 116,52, df = 25,000					
Estadio	1	2	3	4	5
	0,06983	0,05367	8,2022	53,875	3,1502
Huevo		1	1	0	1
Larva 1	1		1	0	1
Larva 3	1	1		0,000001	1
<b>Pupa</b>	0	0	0,000001		0
Adulto	1	1	1	0	

#### ANOVA de un factor: Condición

Effect	Univariate Tests of Significance for dsx NRQ (Spreadsheet2) Sigma-restricted parameterization Effective hypothesis decomposition				
	SS	Degr. of Freedom	MS	F	p
Intercept	5124,93	1	5124,93	9,486144	0,004605
Condición	539,61	1	539,608	0,998803	0,326159
Error	15127,12	28	540,254		

En este caso al analizar todos los estadios en conjunto el resultado no es significativo. Se procede a analizar para cada estadio la diferencia en el factor condición

- Huevos

Effect	Univariate Tests of Significance for Var3 (Spreadsheet71) Sigma-restricted parameterization				
	SS	Degr. of Freedom	MS	F	p

	Effective hypothesis decomposition				
	SS	Degr. of Freedom	MS	F	p
Intercept	0,02926	1	0,02926	89,16252	0,000701
"Var2"	0,000048	1	0,000048	0,14678	0,721126
Error	0,001313	4	0,000328		

- Larva 1

Effect	Univariate Tests of Significance for Var3 (Spreadsheet71) Sigma-restricted parameterization Effective hypothesis decomposition				
	SS	Degr. of Freedom	MS	F	p
Intercept	0,017281	1	0,017281	39,21483	0,003317
"Var2"	0,001121	1	0,001121	2,54312	0,186004
Error	0,001763	4	0,000441		

- Larva 3

Effect	Univariate Tests of Significance for Var3 (Spreadsheet71) Sigma-restricted parameterization Effective hypothesis decomposition				
	SS	Degr. of Freedom	MS	F	p
Intercept	403,6532	1	403,6532	57,06276	0,001646
"Var2"	57,2701	1	57,2701	8,09603	0,046614
Error	28,2954	4	7,0738		

- Pupa

Effect	Univariate Tests of Significance for Var3 (Spreadsheet71) Sigma-restricted parameterization Effective hypothesis decomposition				
	SS	Degr. of Freedom	MS	F	p
Intercept	17415,31	1	17415,31	843,6597	0,000008
"Var2"	2689,44	1	2689,44	130,2859	0,000336
Error	82,57	4	20,64		

- Adulto

	Univariate Tests of Significance for Var3 (Spreadsheet71)
--	---

Effect	Sigma-restricted parameterization Effective hypothesis decomposition				
	SS	Degr. of Freedom	MS	F	p
Intercept	59,5413	1	59,5413	2127,488	0,000001
"Var2"	55,4192	1	55,4192	1980,2	0,000002
Error	0,11195	4	0,02799		

Reactivo	Masa (g) para un litro	Concentración final
Tris base	121, 1	1M
Ácido Bórico	61,8	1M
EDTA (Etilendiamina disódica)	7,4	0.02M
Llevar a volúmen final 100 ml con agua bidestilada libre de nucleasas. Para utilizar diluir a volumen final 1 litro.		

- **Buffer TBE**

- **Geles agarosa**

**0,8%:** se pesaron 0,7g de agarosa y se disolvieron en 60 ml de buffer TBE calentando en microondas hasta observar la solución completamente translúcida.

**1%:** se pesaron 0,9g de agarosa y se disolvieron en 60 ml de buffer TBE calentando en microondas hasta observar la solución completamente translúcida.

**1,5%:** se pesaron 1,2g de agarosa y se disolvieron en 60 ml de buffer TBE calentando en microondas hasta observar la solución completamente translúcida.